AF
2740

PATENT

IN THE UNITED STATES PATENT AND TRADEMARK OFFICE

| | | |
|---|---|---|
| Applicant: | De Francesco, R. *et al.* | |
| Serial No.: | 10/085,476 Case No.: ITR0002PCA | Art Unit: 1652 |
| Filed: | February 27, 2002 | Examiner: Hutson, Richard G. |
| For: | METHOD OF REPRODUCING IN VITRO THE RNA-DEPENDENT RNA POLYMERASE AND TERMINAL NUCLEOTIDYL TRANSFERASE ACTIVITIES ENCODED BY HEPATITIS C VIRUS (HCV) | |

Commissioner for Patents
P.O. Box 1450
Alexandria VA 22313-1450

Response to Communication Re: Appeal

Sir:

Responsive to the Communication Re: Appeal, mailed April 26, 2006, enclosed is an amended copy of the previously submitted Appeal Brief. The enclosed Appeal Brief differs from the previously submitted brief by using the heading "RELATED PROCEEDINGS APPENDIX", and indicating in the Table of Contents that the different listed appendices are attached.

The Communication indicated that applicant's Appeal Brief was not acceptable and the appeal was dismissed. Specifically, the brief was indicated not to contain a heading for (x) Related proceedings appendix.

Applicant discussed the Communication with the examiner in a phone conference May 3, 2006. Applicant indicated that applicant's file copy of the Appeal Brief included a separate page captioned "RELATED PROCEEDINGS" and indicated "None". The "RELATED

---

37 C.F.R. 1.8 Certificate of Mailing

I hereby certify that this correspondence is being deposited with the United States Postal Service as first class mail in an envelope addressed to: Commissioner for Patents P.O. Box 1450 Alexandria VA 22313-1450, on the date appearing below.

MERCK & CO., INC.

By ____________________ Date May 5, 2006

Sheldon O. Heber

PROCEEDINGS" page was submitted after the Evidence Appendix, which ended with U.S. Patent No. 5,981,247.

The examiner indicated that the patent office file copy of the Appeal Brief did not contain U.S. Patent No. 5,981,247 or the separate page captioned "RELATED PROCEEDINGS". The examiner also indicated that the appeal was not dismissed and applicant is being given an opportunity to correct the noted defect.

The enclosed Appeal Brief corresponds to applicant's copy of the previously submitted Appeal Brief, with the above-noted changes to the Table of Contents and use of the heading "RELATED PROCEEDINGS APPENDIX". The enclosed brief includes a copy of U.S. Patent No. 5,981,247.

If any time extensions are needed for the timely filing of the Appeal Brief, applicants petition for such extensions and authorize the charging of deposit account 13-2755 for the appropriate fees. If other fees are due in connection with the enclosed Appeal Brief applicants authorize the charging of deposit account 13-2755 for the appropriate fees.

Respectfully submitted,

By Sheldon O. Heber

Sheldon O. Heber
Reg. No. 38,179
Attorney for Applicant(s)

MERCK & CO., INC.
P.O. Box 2000
Rahway, New Jersey 07065-0907
(732) 594-1958

# BEFORE THE BOARD OF PATENT APPEALS AND INTERFERENCES

Appellant(s): De Francesco, R. *et al.*

Application Number: 10/085,476

Filing Date: February 27, 2002

Title of the Invention: METHOD FOR REPRODUCING IN VITRO THE RNA-DEPENDENT RNA POLYMERASE AND TERMINAL NUCLEOTIDYL TRANSFERASE ACTIVITIES ENCODED BY HEPATITIS C VIRUS (HCV)

Examiner: Huston, Richard G

Art Unit: 1652

# APPEAL BRIEF

37 C.F.R. 1.8 Certificate of Mailing

I hereby certify that this correspondence is being deposited with the United States Postal Service as first class mail in an envelope addressed to: Commissioner for Patents P.O. Box 1450 Alexandria VA 22313-1450, on the date appearing below.

MERCK & CO., INC.

By Sheldon O. Heber

Date May 5, 2006

Sheldon O. Heber

## TABLE OF CONTENTS

## REAL PARTY IN INTEREST

The real parties in interest are Istituto Di Ricerche Di Biologia Molecolare P. Angeletti S.P.A., and Merck & Co., Inc.

## RELATED APPEALS AND INTERFERENCES

There are no pending related appeals and interferences. An Appeal Brief was filed in the parent application U.S. Serial No. 08/952,981. The parent application was allowed upon filing the Appeal Brief and did not go before the Board of Patent Appeals and Interferences. The parent application issued as U.S. Patent No. 6,383,768. A Terminal Disclaimer was filed in the present application with respect to U.S. Patent No. 6,383,768.

## STATUS OF CLAIMS

Claims 12, 14, 17, 18, 22 and 23 stand rejected. Claims 20 and 21 are allowed taking into consideration an after final amendment mailed September 14, 2005. The rejection to claims 12, 14, 17, 18, 22 and 23 is being appealed. The claims involved in the appeal (claims 12, 14, 17, 18, 22 and 23) are provided in the Claims Appendix.

## STATUS OF AMENDMENTS

A final amendment was filed September 14, 2005 addressing objections to claims 20 and 22. The advisory action mailed October 28, 2005 indicated that for the purposes of appeal the amendment would be entered.

## SUMMARY OF CLAIMED SUBJECT MATTER

The present appeal includes two independent claims: claims 12 and 22. Claims 12 and 22 are both directed to a method of identifying a Hepatitis C virus (HCV) RNA-dependent RNA polymerase inhibitor using HCV NS5B, where NS5B was expressed in either a eukaryotic or prokaryotic heterologous system. Reference to a eukaryotic or prokaryotic heterologous system indicates that the employed NS5B was recombinantly expressed using an artificial expression system.

HCV is a virus that infects human liver cells and replicates in the infected cells. During replication in an infected cell, the HCV genome produces a precursor polyprotein. The precursor polyprotein is then cleaved into different proteins, which have different activities.

The present application successfully establishes that NS5B produced in an artificial expression system provides RNA-dependent RNA polymerase activity encoded by HCV. NS5B is part of an HCV region designated NS5. The application further demonstrates that NS5B can be successfully purified to apparent homogeneity and have sufficient activity to be used in a method for identifying a HCV RNA-dependent RNA polymerase inhibitor.

The method of claim 12 employs a HCV NS5B that was expressed in either a eukaryotic or prokaryotic heterologous system and purified to apparent homogeneity. The method involves incubating *in vitro* a composition comprising the purified NS5B, ribonucleotide substrates, an RNA template, and a test compound under conditions suitable to provide RNA-dependent RNA polymerase activity in the absence of the inhibitor; and measuring the ability of the compound to affect RNA-dependent RNA polymerase activity.

Claim 22 is along the same lines as claim 12, but does not indicate the use of NS5B purified to apparent homogeneity. Claim 22 does indicate that HCV NS5B was expressed in either a eukaryotic or prokaryotic heterologous system. Claim 22 is broader than claim 12.

## GROUNDS OF REJECTION TO BE REVIEWED ON APPEAL

I. Claims 12, 14, 17, 18, 22 and 23 stand rejected as allegedly obvious based on Tomei et al. (Journal of Virology 67(7): 4017-4026, July 1993).

## ARGUMENT

I. Claims 12, 14, 17, 18, 22 and 23 are Not Obvious Based on Tomei et al.

Obviousness under 35 U.S.C. § 103 is examined in light of the following factual inquires: (1) the scope and content of the prior art; (2) the differences between the prior art and claims at issue; (3) the level of ordinary skill in the art; and (4) secondary considerations. *Graham v. John Deere*, 383 U.S. 1, 17-18 (1966).

The provided obviousness rejection argues for different modifications to Tomei et al. The argued for modifications are viewed in the context of: (1) whether the prior art would have suggested those of ordinary skill in the art to make the claimed composition or carryout the claimed process; and (2) whether the prior art also reveals that those of ordinary skill in the art would have a reasonable expectation of success in making the composition or carrying out the process. *In re Dow Chemical,* 837 F.2d 469, 473, 5 USPQ2d 1529, 1531 (Fed Cir. 1988).

The suggestion and expectation of success must both be founded in the prior art, not in the applicant's disclosure. *Id.* In determining if ". . . such a suggestion can fairly be gleaned from the prior art, the full field of the invention must be considered; for the person of ordinary skill is charged with knowledge of the entire body of technological literature, including that which might lead away from the claimed invention." *In re Dow Chemical*, 837 F.2d at 473, 5 USPQ2d at 1531-1532.

### A. Claim 22 Distinguishes Tomei et al. by Employing HCV NS5B Expressed in Either a Eukaryotic or Prokaryotic Heterologous System

The method of claim 22 distinguishes Tomei et al. by, for example, employing HCV NS5B expressed in either a eukaryotic or prokaryotic heterologous system to identify a RNA-dependent RNA polymerase inhibitor. Eukaryotic and prokaryotic heterologous systems are artificial expression systems that provide recombinantly expressed NS5B. Claim 22 describes using an *in vitro* composition comprising the recombinantly expressed NS5B under conditions where enzyme activity is produced in the absence of compound and measuring the ability of the compound to affect enzyme activity.

The rejection fails to consider the prior art as a whole, which includes references teaching away from the claimed invention. The prior art expresses doubt as to whether recombinantly expressed HCV NS5B is an authentic HCV protein, the prior art fails to provide data concerning NS5B activity, and the prior art provides evidence of secondary considerations supporting patentability.

The prior art uncertainties concerning the relevance of recombinantly produced NS5B to a naturally occurring HCV protein impacts both: (1) motivation to modify Tomei et al. as suggested in the obviousness rejection; and (2) the likelihood of success in modifying Tomei et al. The uncertainty as to the relevance of recombinantly produced NS5B points away from the skilled artisan being motivated to use recombinantly expressed NS5B in a HCV RNA-dependent RNA polymerase assay looking for polymerase inhibitors. Such uncertainty is also evidence that at the time the invention was made, the skilled artisan would not have a reasonable expectation of success in using NS5B to generate HCV RNA-dependent RNA polymerase activity.

1. Tomei et al. Concerns HCV Polyprotein Processing Using a Recombinant Expression System

Tomei et al. identifies NS3 as a serine protease required for HCV polyprotein process. Tomei et al. does not provide data concerning recombinantly expressed NS5B activity or indicate that NS5B could be used in an assay to look for RNA-dependent RNA polymerase activity inhibitors. Instead of considering the relevance of the observations provided by Tomei et al. in light of the prior art as a whole, the rejection focuses on certain statements provided in Tomei et al. and assumes particular motivations.

The obviousness rejection argues that Tomei et al. should be modified based on different motivations: (1) motivation to incubate NS5B, ribonucleotide substrates and an RNA template to characterize the function and role of protein(s) encoded by the NS5B open reading frame (ORF); (2) motivation to produce NS5B to determine whether proteolytic processing affects NS5B protein product; (3) motivation to vary the RNA templates and primers to characterize RNA-dependent RNA polymerase specific mechanism of action; (4) motivation to add ribonucleotide substrates and RNA template based on the suggestion in Tomei et al. that NS5B may encode a RNA-dependent RNA polymerase; and (5) motivation to identify potential HCV therapeutics

against HCV. (Final Rejection mailed 7/27/05, at pages 3-5, and the Advisory Action mailed 10/28/05 at page 2, first paragraph.)

The rejection argues for a reasonable expectation of success based on Tomei et al. and the high level of skill in the art. In discussing a reasonable expectation of success in modifying Tomei et al. to obtain the claimed assay, the rejection refers to Tomei et al. suggesting that the NS5B open reading frame encodes a RNA-dependent RNA polymerase, the high level of skill in the art, the HCV genome being processed in a similar manner as flaviviruses and pestiviruses, and the hydropathy profile of the HCV polyprotein being similar to that of the flavivirus. (See Final Rejection mailed 7/27/05, at pages 3-5, and the Advisory Action mailed 10/28/05 at page 2, first paragraph.)

2. <u>The Significant Uncertainty Expressed in the Prior Art as a Whole Concerning Whether Recombinantly Expressed NS5B is an Authentic HCV Protein or a Recombinant Expression Artifact Points Against both Motivation and a Reasonable Expectation of Success in Modifying Tomei et al. to Obtain the Claimed Assay</u>

Prior to the present application there was significant uncertainty concerning the relevance of HCV NS5B produced in recombinant expression systems, such as that employed by Tomei et al., to a naturally produced HCV protein product. NS5B is not directly produced from an open reading frame encoding the protein. The HCV open reading frame encodes a polyprotein. NS5B is produced from proteolytic processing of polyprotein.

Motivations asserted in the obviousness rejection such as characterizing the function and role of NS5B and determining whether proteolytic processing affects NS5B, reflect the uncertainties in the art. Such motivations amount to an invitation to study HCV processing and do not provide motivation, or a reasonable expectation of success, to screen for inhibitors using recombinantly expressed NS5B. The high level of skill in the art relates to the ability of the skilled artisan to perform certain activities and does not resolve the scientific uncertainties existing in the prior art.

The uncertainties concerning the relevance of recombinantly produced NS5B to authentic HCV products are reflected in the differences between published results obtained from HCV

infected liver cell versus recombinantly expressed HCV polyprotein. The uncertainties are also noted in cautionary language used in publications concerning recombinantly produced NS5B.

Tsutsumi et al. (Hepatology 19(2), 265-272, 1994) observed that HCV in infected liver cells did not produce a protein with approximately the same molecular weight as NS5B:

> Grakoui et al. (15) reported that two proteins were derived from the HCV-NS5 region: NS5A (58kD) and **C-terminal NS5B (66 to 68 kD)**, when a cDNA encompassing the long open reading frame was used in vaccinia virus transient-expression assay. The NS5B protein was predicted to contain the RNA-dependent RNA polymerase activity on the basis of the presence of the characteristic Gly-Asp-Asp and surrounding conserved motifs. Although bacterially expressed HCV-NS5 peptide fragment was used for a part of NS5B protein in this study, the molecular size of **HCV-NS5-related antigen detected in human liver was 86 kD** and thus slightly larger than that of NS5B. This discrepancy may have resulted in different host cells, which were cultured mammalian cells in Grakoui's study (15) and were human liver cells in this study. **Furthermore, we observed the products derived from native HCV, whereas Grakoui et al. (15) observed the polypeptide expressed from HCV cDNA in vaccinia virus**. [Emphasis added.]

(Tsutsumi et al. starting at page 269, first column, second paragraph to page 270, first column.)

The magnitude of the different molecular weights for the infected human liver and recombinantly processed HCV NS5 regions point to different proteins and not minor variations. The 86 kD molecular weight noted by Tsutsumi et al. is 21 kD or 25 % more than the Tomei et al. NS5B weight of 65 kD; and 18 kD or 21 % more than the upper 68 kD attributed to Grakoui et al.

Tomei et al. mentions recombinantly produced NS5B, points out that the HCV NS5 region is processed differently than flavivirus NS5B and merely speculates that NS5B may act as a viral replicase:

> The NS5 region of the HCV polyprotein is cleaved into two smaller products of 47 and 65 kDa; **the processing of this region therefore differs from that of flavivirus NS5**, which is released from the polyprotein precursor as a single protein of 110 kDa. The GDD consensus sequence characteristic of RNA-dependent RNA polymerases is located in NS5b (residues 2736 to 2738), indicating that this protein **may** act as a viral replicase during HCV-specific RNA synthesis (17). However, NS5a could also have a function in the replication of the viral genome, acting as a component of the replication complex involved in the reaction. [Emphasis added.]

(Tomei et al. at page 4024, column 1, fifth paragraph.)

In considering the relevance of recombinantly produced NS5B, Tomei et al. in its concluding paragraph points out that recombinantly produced NS5B may not correspond to a naturally produced product:

> **It is clear, however, that the results obtained with this transient expression system may not faithfully reproduce the proteolytic events which take place during HCV infection.** It is possible that the level of protein expression obtained in this system may be much higher than normal, affecting important equilibria between precursors and proteases, which in turn may regulate HCV replication and protein synthesis. [Emphasis added.]

(Tomei et al., at page 4025, first column, last paragraph.)

Grakoui et al. (Journal of Virology 67(3), 1385-1395, 1993) in its concluding paragraph expresses a similar concern concerning the relevance of recombinant NS5B:

> The experiments reported here have given us a preliminary picture of HCV polyprotein organization and processing. However, **this view is far from complete, and additional studies are needed to define polyprotein cleavage sites** and the responsible proteinases and to verify that the products observed in these expression studies are similar to those produced in authentic HCV infections. Such information should prove valuable for expression and characterization of HCV-encoded enzymes as potential targets for antiviral therapy and will allow future studies . . . [Emphasis added.]

(Grakoui et al. at page 1393, first column, third paragraph.)

The present application demonstrates that recombinant NS5B provides for an HCV RNA-dependent RNA polymerase. Neither Tomei et al., Grakoui et al., or Tsutsumi et al. provide results demonstrating that an observed protein provides HCV RNA-dependent RNA polymerase activity. Thus, the present application, not the prior art, resolves the scientific uncertainties concerning the relevance of NS5B to providing RNA-dependent RNA polymerase activity.

3. <u>Apparent Failure, Difficulty Encountered by Others and Long-Felt Need Further Illustrate the Non-Obvioussness of the Claimed Assay</u>

Additional considerations illustrating the inventive nature of the claimed assay include: (1) apparent failure and difficulty encountered by others in demonstrating the HCV region responsible for RNA-dependent RNA polymerase; and (2) a long-felt need for an HCV RNA-dependent RNA polymerase assay to look for polymerase inhibitors. Apparent failure and

difficulty encountered by others is evident based U.S. Patent No. 5,981,247[1]. The long-felt need is apparent based on the importance of HCV and the time delay between published speculation concerning the HCV RNA-dependent RNA polymerase and the present application.

U.S. Patent No. 5,981,247 was filed September 27, 1996, and claims priority to a provisional application dated September 27, 1995. The provisional application date is about four months after the priority date for the present application. According to U.S. Patent No. 5,981,247, in the Background of the Invention:

> The non-structural protein designated 5B (NS5B) has been shown to have an amino-terminal sequence SMSY (Ser-Met-Ser-Tyr). The NS5B region encodes a 68 kd protein (p68) which contains an internal GDD (Gly-Asp-Asp) motif found in RNA-dependent RNA polymerases of other RNA viruses (Koonin, E. V. (1991) *J. Gen. Virol.* 72:2197-2206). **However, no polymerase activity has been detected for HCV p68.** In fact, the question has been raised that the 5B protein (p68) alone does not encode an active RNA-dependent RNA polymerase enzyme and that another subunit, possibly the NS5A gene product, is essential to catalytic activity. **Prior attempts by the inventors and others to express the NS5B coding region as a fusion protein, using existing expression systems that facilitate purification of the fusion product and specific cleavage have failed to yield any active polymerase.** [Emphasis added.]

(U.S. Patent No. 5,981,247, at column 1, line 59 to column 2, line 7.)

The long felt-need is evident based on the medical importance of HCV, the desirability of an assay to screen for a HCV RNA-dependent RNA polymerase inhibitor, and the time difference between prior art speculations concerning HCV RNA-dependent RNA polymerase and applicant's priority application. Speculations concerning the HCV protein responsible for RNA-dependent RNA polymerase are noted at least as early as 1990. (*E.g.,* Miller et al., Proc. Natl. Acad. Sci. USA, March 1990, 87(6), 2057-2061, at page 2061, first column, third paragraph). The present application has a priority date of May 25, 1995, which is more than five years after Miller et al. was published.

---

[1] U.S. Patent No. 5,981,247 corresponds to WO 97/12033. WO 97/12033, referenced by applicants during prosecution, was mailed to the Patent Office as part of a supplemental information disclosure statement (IDS) on April 26, 2005. According to the public PAIR and applicants records the IDS was received. However, applicants have not received an initialed version of the IDS. U.S. Patent No. 5,981,247 was previously made of record and is referenced in the present Appeal Brief.

4. <u>The Provided Rejection Improperly Failed to Consider Prior Art Uncertainty as to Whether Recombinantly Expressed NS5B is an Authentic HCV Protein or a Recombinant Expression Artifact and Failed to Consider Secondary Considerations</u>

Applicants' arguments concerning prior art uncertainty on the importance of recombinantly expressed NS5B and secondary considerations appear to have been given no weight by the examiner. The examiner dismisses applicants arguments concerning prior art uncertainty on basis that applicants have not presented sufficient evidence that NS5B is an expression artifact. (Final Rejection mailed 7/27/05, at page 6, third paragraph; and Advisory Action mailed 10/28/05, at page 2 sixth paragraph.) The examiner dismisses applicants' arguments concerning secondary consideration on the basis that such arguments are apparently not directed to the rejection of record as it applies to the claims. (Final Rejection mailed 7/27/05, at page 7, third paragraph; and Advisory Action mailed 10/28/05, at page 2, seventh paragraph.)

The basis asserted by the examiner for dismissing applicants' arguments concerning prior art uncertainty and secondary consideration are improper. Applicants' arguments concerning prior art uncertainty goes to the state of art when the invention was made. Applicants are not arguing whether or not NS5B is in fact an expression artifact. In hindsight, based on the present application it is known that NS5B provides for RNA-dependent RNA polymerase activity.

The uncertainty in the art at the time the invention was made concerning the relevance of recombinantly expressed NS5B points away from motivation to use NS5B in an assay to identify RNA-dependent RNA polymerase inhibitors and points away from a reasonable expectation of success. That the inventors were ultimately successful is irrelevant to whether the skilled artisan would have reasonably expected success at the time the invention was made. *Life Technologies Inc. v. Clontech Laboratories Inc.*, 224 F.3d 1320, 1326, 56 USPQ2d 1186, 1191 (Fed. Cir. 2000).

The difficulty and apparent failure encountered by others in obtaining an active NS5B is directly relevant to claims. Claim 22 indicates the production of NS5B RNA-dependent RNA polymerase activity. Failure of other to produce such activity provides evidence of non-obviousness.

## B. Claim 23 Further Distinguishes Tomei et al. by Measuring Primer Independent RNA-Dependent RNA Polymerase Activity

Claim 23, which depends from claim 22, further distinguishes Tomei et al. by measuring primer independent RNA-dependent RNA polymerase activity. The rejection appears to argue that primer independent RNA-dependent RNA polymerase activity would be apparent based on characterization of NS5B RNA-dependent RNA polymerase activity.

The rejection amounts to an invitation to experiment to characterize NS5B activity. The prior art fails to even demonstrate that NS5B provides for RNA-dependent RNA polymerase activity. Absent knowing that NS5B provides for RNA-dependent RNA polymerase activity the skilled artisan would not be motivated to further characterize the enzyme, or set up an assay to look for inhibitors by measuring primer independent RNA-dependent RNA polymerase activity.

## C. Claims 12, 17 and 18 Further Distinguishes Tomei et al. by Employing NS5B Purified to Apparent Homogeneity

Claim 12 is along the lines of claim 22, but indicates that NS5B is purified to apparent homogeneity. Claims 17 and 18 depend from clam 12, and for the purposes of the rejection are argued with claim 12.

Reference to purified to apparent homogeneity further distinguishes Tomei et al. by indicating a very high degree of purity. Claim 12 describes using such highly purified protein to provide for RNA-dependent RNA polymerase activity.

The Patent Office bears the initial burden of presenting a *prima facie* case of unpatentability. *In re Oetiker,* 977 F.2d 1443, 1445, 24 USPQ2d 1443, 1444 (Fed. Cir. 1992). The rejection merely refers to high degree of skill in the art of protein purification.

In addition, U.S. Patent No. 5,981,247 and Chung et al. (Hepatology, 16(4), 1992) point to prior art difficulties in obtaining purified HCV RNA-dependant RNA polymerase. U.S. Patent No. 5,981,247 refers to prior unsuccessful attempts to purify enzymatically active NS5B. (U.S. Patent No. 5,981,247, at column 1, line 59 to column 2, line 7, noted in applicants Argument Section I.A.3. *supra.*)

Chung et al. is an abstract mentioning attempts to obtain HCV RNA-dependent RNA polymerase from liver tissue. Chung et al. references activity obtained with partially purified extracts. Reference to only partially purified activity is consistent with failed attempts to obtain

a purified product. Failed purification attempts are evident based on the use of different chromatographic techniques and the desirability to obtain a purified enzyme for further study.

D. Claim 14 Further Distinguishes Tomei et al. by Measuring Primer Independent RNA-dependent RNA Polymerase Activity.

Claim 14, which depends from claim 12, further distinguishes Tomei et al. by measuring primer independent RNA-dependent RNA polymerase activity. The rejection appears to argue that primer independent RNA-dependent RNA polymerase activity would be apparent based on characterization of NS5B RNA-dependent RNA polymerase activity.

The rejection amounts to an invitation to experiment to characterize NS5B activity. The prior art fails to even demonstrate that NS5B provides for RNA-dependent RNA polymerase activity. Absent knowing that NS5B provides for RNA-dependent RNA polymerase activity the skilled artisan would not be motivated to further characterize the enzyme, or set up an assay to look for inhibitors by measuring primer independent RNA-dependent RNA polymerase activity.

## CONCLUSION

Appellants request that the Board of Patent Appeals and Interferences reverse the outstanding rejections of claims 12, 14, 17, 18, 22 and 23.

Please charge deposit account 13-2755 for fees due in connection with this Appeal Brief. If any time extensions are needed for the timely filing of the present Appeal Brief, Appellants petition for such extensions and authorize the charging of deposit account 13-2755 for the appropriate fees.

Respectfully submitted,

By ______________

Sheldon O. Heber
Reg. No. 38,179
Attorney for Applicant(s)

Merck & Co., Inc.
RY60-30
P.O. Box 2000
Rahway, NJ 07065-0907
(732) 594-1958

## CLAIMS APPENDIX

Claim 12. A method for identifying a HCV RNA-dependent RNA polymerase inhibitor comprising:

(a) incubating *in vitro* a composition comprising a purified HCV NS5B recombinant protein, ribonucleotide substrates, an RNA template, and a test compound, under conditions suitable to produce NS5B RNA-dependent RNA polymerase activity in the absence of said compound, wherein said recombinant protein was expressed in either a eukaryotic or prokaryotic heterologous system and purified to apparent homogeneity; and

(b) measuring the ability of said compound to affect said NS5B RNA-dependent RNA polymerase activity.

Claim 14. The method of claim 12, wherein said method measures primer independent RNA-dependent RNA polymerase activity.

Claim 17. The method of 12, wherein said NS5B has the amino acid sequence of SEQ ID NO:1.

Claim 18. The method of claim 12, wherein said NS5B is produced from a NS2-NS3-NS4-NS5 polyprotein by means of multiple proteolytic events that occur in an organism expressing nucleic acid encoding said NS2-NS3-NS4-NS5 polyprotein, followed by purification of said NS5B.

Claim 22. A method for identifying a HCV RNA-dependent RNA polymerase inhibitor comprising:

(a) incubating *in vitro* a composition comprising HCV NS5B, ribonucleotide substrates, an RNA template, and a test compound, under conditions suitable to produce NS5B RNA-dependent RNA polymerase activity in the absence of said test compound, wherein said HCV NS5B was expressed in either a eukaryotic or prokaryotic heterologous system; and

(b) measuring the ability of said test compound to affect said NS5B RNA-dependent RNA polymerase activity.

Claim 23. The method of claim 22, wherein said method measures primer independent RNA-dependent RNA polymerase activity.

## EVIDENCE APPENDIX

A list of the references in support of applicants' arguments and the date the references were made of record by the examiner is provided below. A copy of each reference is enclosed.

Chung et al. (Hepatology, 16(4), 1992); made of record January 20, 2005.

Grakoui et al. (Journal of Virology 67(3), 1385-1395, 1993); made of record January 20, 2004.

Miller et al., (Proc. Natl. Acad. Sci. USA, March 1990, 87(6), 2057-2061; made of record February 20, 2004.

Tomei et al. (Journal of Virology 67(7): 4017-4026, July 1993); made of record February 20, 2004.

Tsutsumi et al. (Hepatology 19(2), 265-272, 1994); made of record January 20, 2005.

U.S. Patent No. 5,981,247; made of record February 20, 2004.

349 **DETECTION OF HBV DNA BY LIGASE CHAIN REACTION (LCR), A HIGHLY SENSITIVE AND QUANTIFIABLE DNA PROBE ASSAY**

H. Hampl[1], B. Goergen[2], B. Grimm[1], H. Witteler[1], U. Spies[1], R. Decker[1], K.-H. Meyer zum Bueschenfelde[2], R. Sutherland[1], and G. Gerken[2].
[1]Abbott GmbH Diagnostika, R&D Europe, 62 Wiesbaden-Delkenheim, FRG
[2]Med. & Poliklinik, Johannes Gutenberg-Universität, 65 Mainz, FRG

*Introduction:* Detection and quantitation of HBV DNA is very important for monitoring success of interferon therapy of patients with chronic active hepatitis. Currently available commercial tests detecting HBV DNA without target amplification have detection limits of around 2-10pg DNA/ml sample. Assays including amplification of target (Polymerase Chain Reaction, PCR) have shown the clinical need to monitor HBV DNA levels below 2pg/ml. However PCR is not routinely available and technically difficult to perform. We report here a new HBV DNA detection assay based on a semi-quantitative LCR including automated detection with sensitivity equivalent to PCR.
*Materials and Methods:* 20 serum samples from different groups of patients with hepatitis B virus infection (HBV carriers with high viremia (HBV DNA and HBeAg positive), low viremia (anti-HBe and HBV DNA PCR positive) and asymptomatic HBV carriers (anti-HBe positive, PCR negative)) as well as HBV DNA dilution panels were investigated using LCR. HBV-DNA was assayed by solution hybridization assay (Abbott) and by PCR as earlier described (Gerken et al, 1991, Hepatology 13, p158-166). LCR assay for HBV DNA (Hampl et al, 1991, in "PCR-Topics" Springer Verlag, p15-22) was done using a set of 4 probes which originate from a conserved genome region of the HBsAg. Total length of each 2 adjacent probes is 48 bases. After 45 cycles in a thermocycler, analysis was done in the IMxTM-detection system (Fiore 1988, Chem. 34, 1726-1732).
*Results:* LCR assay showed sensitivity of about 10fg HBV-DNA per ml sample, equivalent to about 2000 genome copies/ml. Values were linear from 50-800fg/ml with an intra assay variation of 10%. With a 5ul sample this is equivalent to a detection limit of 10 molecules HBV DNA. LCR assay was able to detect quantitatively HBV-DNA in 16/16 PCR reactive samples. All samples (n=4) positive in solution hybridization assay were found positive in LCR assay. When samples were HBeAg positive, the HBV DNA LCR signal was highly positive in all (n=5) instances. One sample (negative in PCR, HBsAg positive) was found to be weak positive in LCR.
*Conclusion:* A new semi-quantitative molecular assay for HBV-DNA in serum based on LCR technology was developed. Within the limits of this study, sensitivity of the LCR assay is equivalent to PCR with the additional advantage of an automated detection system (Abbott IMx$^{TM}$) analysing 24 samples within 30 minutes.

350 **IDENTIFICATION AND CHARACTERIZATION OF A HEPATITIS C VIRUS-SPECIFIC RNA-DEPENDENT RNA POLYMERASE ACTIVITY FROM EXTRACTS OF INFECTED LIVER TISSUE.**

*RT Chung and LM Kaplan*. Gastrointestinal Unit, Massachusetts General Hospital, Boston, MA

Hepatitis C virus (HCV) is a positive-stranded RNA virus whose genomic organization closely resembles that of the flaviviruses and pestiviruses, which replicate their genomes by direct RNA-to-RNA transcription. Taken together with the lack of evidence for a DNA intermediate in the HCV lifecycle, this genetic similarity suggests that the HCV genome is replicated by means of a viral RNA-dependent RNA polymerase (RdRp) activity. Using a partially double-stranded RNA as primer-template, we have developed an assay to detect RdRp activity in HCV-infected liver tissue. We prepared cellular extracts from fresh native liver tissue obtained from two HCV-infected and two uninfected recipients of orthotopic liver transplants. For each preparation, tissue homogenates and various subfractions were incubated in an RNA polymerase assay mixture containing buffer, divalent cations and all four ribonucleoside triphosphates (including $^3$H-UTP), in the presence of a synthetic RNA primer-template containing sequences from the 5'-untranslated region of the HCV genome. Reactions were incubated at 37°C and polymerase activity measured as incorporation of $^3$H-UMP into acid-precipitable material.

We detected RdRp activity in extracts of HCV-infected liver tissue that was dependent on the addition of exogenous primer-template RNA. In contrast, similar activity was not detectable in liver extracts from uninfected individuals. The HCV infection-specific activity is inactivated by heat, is time- and dose-dependent, requires divalent cations, and is resistant to the *DNA-dependent* RNA polymerase inhibitors actinomycin D and α-amanitin. The activity is dependent on the presence of all four ribonucleoside triphosphates, as well as both template and primer RNA strands, indicating that the incorporation reflects the elongation activity of an RNA-dependent RNA polymerase. Digestion with a variety of proteases and nucleases revealed that the product is a double-stranded RNA species. Denaturing agarose gel electrophoresis demonstrated incorporation of radiolabeled nucleotides into full-sized, template-length product after incubation for 30 minutes. We have partially purified the enzymatic activity using a variety of chromatographic separations. The development of an assay for HCV RdRP and the isolation of complexes containing replicase activity should be a useful step in the elucidation of the mechanisms of HCV RNA replication and the identification of compounds that can interrupt the viral replicative cycle.

351 **HCV UNDERGOES EXTENSIVE MUTATIONAL CHANGE IN NS5 REGION IN ASSOCIATION WITH RELAPSE/BREAKTHROUGH FOLLOWING ALPHA-INTERFERON THERAPY.**

Sydney D. Finkelstein, Raoulf Sayegh, Sonia Uchman, Steven Christensen, Patricia Swalsky, Departments of Pathology and Gastroenterology, Rhode Island Hospital, Brown University, Providence, RI

The factors determining interferon responsiveness in the treatment of HCV infection likely involve both virologic and host immunologic mechanisms. To evaluate virus related factors, we reverse transcribed, amplified by single stage (35 cycles) PCR, and directly sequenced HCV RNA from the serum of 15 patients prior to and 6 patients in relapse/breakthrough following alpha-interferon therapy. Four regions of the virus were studied corresponding to the 5' nontranslated, NS3, NS4, and NS5 regions. Nucleotide base sequences and resulting aminoacid (aa) sequences were compared between patients and to the prototype US strain.

Direct sequencing demonstrated similarity to prototype virus confirming specific amplification. In no instance was the HCV gene sequence identical between different patients. Alterations resulting in nucleotide and aa substitutions were found in all cases and for all regions occurring in a nonrandom manner at predictable points of mutability. Insertions or deletions were not identified. The mutational rate was greatest for the NS5 region which proved the most informative. Untreated patients manifested a low rate of NS5 region mutation, averaging 12 per 100 bases and 4 per 100 aa, with some patients showing a higher level of mutability. Patients relapsing following interferon treatment showed a striking degree of NS5 region mutation, averaging 32 per 100 bases and 28 per 100 aa. Nonrandom mutability seen following therapy involved and extended sites already shown to be undergoing change in untreated infection. The evidence suggests that mutability is a fundamental property of the virus, active during untreated disease and especially rapid in relationship to interferon therapy. Mutation of the NS5 region may be associated with treatment failure and its detection may be predictive of ultimate therapeutic responsiveness. It is proposed that those patients in whom the virus has already mutated significantly during untreated disease may represent the subset likely to prove less responsive to standard therapy.

352 **LYMPHOCYTE SUBSETS AND PROLIFERATIVE RESPONSES TO RECOMBINANT HCV ANTIGENS IN PATIENTS WITH CHRONIC HEPATITIS C.**

H. Schupper, P. Hayashi, J. Zeldis, UC Davis Medical Center; T. Paglieroni, S. Aceituno, P. Holland, Sacramento Medical Foundation Center for Blood Research, J. Scheffel, Abbott Laboratories.

**Purpose:** To determine proliferative responses of lymphocyte subsets to recombinant HCV (rHCV) antigens. **Methods:** Lymphocytes from 16 HCV chronic hepatitis patients were examined. All 16 patients were seropositive by enzyme immunoassay (EIA) and 4-antigen recombinant immunoblot assay (RIBA II). 7 healthy seronegative persons were tested as controls. Lymphocyte subsets were determined by dual color flow cytometry. After a 5 day incubation with optimal amounts of recombinant HCV antigen, proliferative responses were measured by staining lymphocytes with propidium iodide then quantitating the % S phase cells by flow cytometry (FACScan, Cellfit software). **Results:** 7 patients had an increase in the percentage of cytotoxic non-MHC-restricted T cells. 2 patients had decreased CD4/CD8 ratios (0.6). Otherwise the numbers of CD19 B cells, CD3, CD4, or CD8 T cells and NK cells in patients did not differ from controls. In vitro proliferative responses to PHA were decreased in 3 of 16 patients.

Lymphocyte Responses to rHCV antigens

| rHCV Ag | #pts w > 4% S phase/ #pts tested | #ctrls w > 4% S/ #ctrls tested |
|---|---|---|
| CKS E.coli ctrl | 1/12 | 1/7 |
| SOD yeast ctrl | 0/12 | 0/7 |
| c22 SOD | 11/11 | 0/7 |
| c100-3 SOD | 9/11 | 0/7 |
| CKS-33c (NS 3) | 6/10 | 0/7 |
| CKS-core | 5/8 | 0/7 |
| CKS-EF (NS 5) | 0/10 | 0/7 |

When antibody was detected on RIBA, there was a lymphoproliferation in response to the relevant HCV antigen. **Conclusions:** 1) there is an increase in an unusual subset of T cells in HCV infected individuals; 2) cellular immune response of PMC T cells parallels humoral immunity.

JOURNAL OF VIROLOGY, Mar. 1993, p. 1385–1395 Vol. 67, No. 3
0022-538X/93/031385-11$02.00/0


# Expression and Identification of Hepatitis C Virus Polyprotein Cleavage Products

ARASH GRAKOUI,[1] CZESLAW WYCHOWSKI,[2]† CHAO LIN,[1] STEPHEN M. FEINSTONE,[2] AND CHARLES M. RICE[1]*

*Department of Molecular Microbiology, Washington University School of Medicine, Box 8230, 660 South Euclid Avenue, St. Louis, Missouri 63110-1093,[1] and Division of Virology, CBER, Food and Drug Administration, Bethesda, Maryland 20892[2]*

Received 12 October 1992/Accepted 13 November 1992

**Hepatitis C virus (HCV) is the major cause of transfusion-acquired non-A, non-B hepatitis. HCV is an enveloped positive-sense RNA virus which has been classified as a new genus in the flavivirus family. Like the other two genera in this family, the flaviviruses and the pestiviruses, HCV polypeptides appear to be produced by translation of a long open reading frame and subsequent proteolytic processing of this polyprotein. In this study, a cDNA clone encompassing the long open reading frame of the HCV H strain (3,011 amino acid residues) has been assembled and sequenced. This clone and various truncated derivatives were used in vaccinia virus transient-expression assays to map HCV-encoded polypeptides and to study HCV polyprotein processing. HCV polyproteins and cleavage products were identified by using convalescent human sera and a panel of region-specific polyclonal rabbit antisera. Similar results were obtained for several mammalian cell lines examined, including the human HepG2 hepatoma line. The data indicate that at least nine polypeptides are produced by cleavage of the HCV H strain polyprotein. Putative structural proteins, located in the N-terminal one-fourth of the polyprotein, include the capsid protein C (21 kDa) followed by two possible virion envelope proteins, E1 (31 kDa) and E2 (70 kDa), which are heavily modified by N-linked glycosylation. The remainder of the polyprotein probably encodes nonstructural proteins including NS2 (23 kDa), NS3 (70 kDa), NS4A (8 kDa), NS4B (27 kDa), NS5A (58 kDa), and NS5B (68 kDa). An 82- to 88-kDa glycoprotein which reacted with both E2 and NS2-specific HCV antisera was also identified (called E2-NS2). Preliminary results suggest that a fraction of E1 is associated with E2 and E2-NS2 via disulfide linkages.**

Prospective and retrospective serologic studies indicate that even before the implementation of screening tests for hepatitis B surface antigen, non-B hepatitis accounted for the majority of transfusion-associated hepatitis in the United States (21, 22, 54). Recently, the major etiologic agent of non-A, non-B hepatitis (NANBH), hepatitis C virus (HCV), was cloned and sequenced (7, 32). This breakthrough has led to the development of immunological and nucleic acid-based methods for detecting HCV infection. HCV infection can result in various clinical outcomes, including acute hepatitis, chronic hepatitis, cirrhosis, or the establishment of an asymptomatic carrier state which may persist for life (for a review, see reference 31). Recent studies have also uncovered a strong association between chronic HCV infection and the development of hepatocellular carcinoma (12, 13, 61).

Since the initial molecular cloning of this agent and implementation of first-generation diagnostics, sequence data for a number of independent HCV isolates have been reported, immunoassays for detection of antibody have been improved, and our knowledge of HCV molecular biology is advancing rapidly (for a review, see reference 33). On the basis of their genome organizations and virion properties, HCV (9, 33), the pestiviruses (11), and the flaviviruses (4) have been classified as three genera in the family *Flaviviridae* (19). Properties shared by these three groups include a lipid envelope, conferring sensitivity to organic solvents, and a single-stranded, positive-polarity RNA genome containing a long open reading frame (ORF) which encodes the viral polypeptides. Polyproteins encoded by the HCV, flavivirus, and pestivirus ORFs are ~3,000, ~3,400, and ~4,000 amino acids long, respectively. The structural proteins are located in the N-terminal portion of the polyprotein and are followed by the putative nonstructural replicase components. Mature proteins, at least as shown for the flaviviruses and pestiviruses (for a review, see references 10 and 58), are produced by a combination of host and viral proteinases located in both the cytosol and the subcellular vesicular compartments.

Although the cleavage products and proteolytic processing schemes of the flaviviruses and pestiviruses have been extensively characterized, similar information has been reported only for the structural protein coding region of HCV (30). In this report, a hybrid vaccinia virus-T7 transient expression system (16, 20, 48) has been used to study processing of the entire HCV ORF. HCV-specific cleavage products were identified by using a collection of region-specific polyclonal rabbit antisera. These results provide a preliminary picture of HCV processing and a map of the polyprotein cleavage products.

## MATERIALS AND METHODS

**Cell cultures and virus growth.** The BHK-21 and CV-1 cell lines were obtained from the American Type Culture Collection (ATCC), Rockville, Md., the BSC-40 cell line (3) was obtained from D. Hruby (Oregon State University), and the A16 subclone of the human hepatoma HepG2 cell line (ATCC) was generously provided by Alan Schwartz (Wash-

* Corresponding author.
† Present address: Unité de Virologie Moléculaire, Institut Pasteur, 75724 Paris Cedex 15, France.

TABLE 1. HCV immunogens expressed in *E. coli*

| Construct[a] | HCV restriction site[b] | Vector site[c] | Rabbit no. | Specificity |
|---|---|---|---|---|
| **pET-3xa/HCV 1-142** | ***ApaLI*** **(335),** ***NarI*** **(762)** | *Bam*HI[d] | WU120 | C |
| **pET-3xa/HCV 236-382** | ***XhoI*** **(1046),** ***SalI*** **(1482)** | *Bam*HI[d] | WU122 | E1 |
| **pMALc/HCV 393-670** | *Nae*I (1515), ***BspMI*** (2357) | *Stu*I | WU105 | E2 |
| pET-8c/HCV 936-1032 | *Nco*I[e], *Spl*I (3433) | *Nco*I, ***BamHI*** | WU107 | NS2 |
| **pMALc/HCV 1039-1207** | ***StuI*** (3452), ***BamHI***[e] | ***EcoRI*** | WU110 | NS3 |
| pET-8c/HCV 936-1207 | *Nco*I[e], *Bam*HI[e] | *Nco*I, *Bam*HI | WU43 | NS2, NS3 |
| **pMALc/HCV 1240-1488** | ***Eco0109*** (4057), *Bam*HI[e] | *Stu*I, *Bam*HI | WU117 | NS3 |
| **pMALc/HCV 1651-1976** | ***EagI*** (5290), ***EcoNI*** (6265) | ***EcoRI*** | WU111 | NS4A, NS4B |
| **pET-3xa/HCV 1977-2313** | ***EcoNI*** (6265), ***NcoI*** (7274) | *Bam*HI[d] | WU123 | NS5A |
| **pMALc/HCV 2312-2623** | ***NcoI*** (7274), *Eco*RI (8205) | *Stu*I, *Eco*RI | WU113 | NS5A, NS5B |
| **pMALc/HCV 2622-2872** | *Eco*RI (8205), *Xba*I[f] | *Eco*RI, *Xba*I | WU115 | NS5B |

[a] Constructs producing fusion proteins are shown in bold type. The numbers refer to the amino acid sequence of the HCV ORF which is included in each construct.

[b] Restriction sites in the HCV cDNA used for the plasmid constructs. Nucleotide numbers in parentheses refer to the full-length HCV H-strain sequence (14), assuming that the 5' noncoding region of HCV H and HCV type 1 (8, 28) are the same length (341 nucleotides). Restriction sites shown in bold type indicate that protruding ends were treated with either the Klenow fragment of DNA polymerase I or T4 DNA polymerase before ligation to produce blunt ends.

[c] Restriction sites in plasmid vectors used for cloning. Sites shown in bold type indicate that protruding ends were treated with either the Klenow fragment of DNA polymerase I or T4 DNA polymerase, prior to ligation to produce blunt ends.

[d] In the case of pET-3xa/HCV 1-142, pET-3xa/HCV 236-382, and pET-3xa/HCV 1977-2313, the indicated HCV cDNA fragments were first cloned into the *Stu*I site of pMALc. The *Bam*HI fragments from these pMALc constructs (containing the properly oriented HCV insert) were then subcloned into *Bam*HI-digested pET-3xa.

[e] For constructs beginning with amino acid residue 936 or ending with residue 1207 or 1488, the restriction sites used for cloning, *Nco*I and *Bam*HI, respectively, were derived by PCR.

[f] This *Xba*I site is in the pGEM-3Zf(+) HCV(H) 17-8958 polylinker.

ington University, St. Louis, Mo.). Cell monolayers were grown in Eagle's minimal essential medium (MEM) supplemented with 2 mM L-glutamine, nonessential amino acids, penicillin, streptomycin, and 10% fetal bovine serum (FBS).

Stocks of vTF7-3, a vaccinia virus recombinant expressing the T7 DNA-dependent RNA polymerase (20), and various vaccinia virus-HCV recombinants were grown in BSC-40 monolayers and partially purified (34), and titers of infectious progeny were determined by plaque assay on BSC-40 cells (34).

**HCV cloning and sequence analysis.** Cloning and sequence analysis of the HCV Hutchinson (H) strain (18) are only briefly described here. The HCV H strain, a human isolate from an American with posttransfusion NANBH, was passaged twice in chimpanzees. Both of these animals developed elevated serum alanine amino transferase levels and acute hepatitis. Liver tissue from the second chimpanzee passage was used for preparation of crude RNA suitable for cDNA synthesis, nested polymerase chain reaction (PCR) amplification (60, 72), and cloning. Synthetic oligonucleotide primers for amplification of specific regions of the HCV genome were originally synthesized on the basis of the published HCV type 1 cDNA sequence (8 [and references therein]). PCR-amplified cDNA was cloned into bacterial plasmid vectors, and several independent clones were isolated and used for sequence analysis, expression studies, and reconstruction of longer cDNA clones. Utilizing partial sequence data and restriction enzyme mapping, a clone containing the entire ORF, called pGEM-3Zf(+) HCV(H) 17-9389F (see below), has been assembled. The clone has been completely sequenced (14) by the Sanger method (62) with a set of synthetic oligonucleotide primers whose sequences were based on preliminary H strain sequence data. The sequence of this clone is colinear and is >98.5% homologous (at the nucleotide level) to recently published full-length (35) and partial (50) HCV H strain sequences.

**Bacterial expression constructs.** Constructs were made by using standard methodology (62), and regions amplified by PCR (60) were verified by sequence analysis (62). *Escherichia coli* expression systems included the pET-3x series, which produce N-terminal fusions with the T7 gene *10* product (66), pMALc derivatives, which produce N-terminal fusions with *E. coli* maltose-binding protein (New England Biolabs), and pET-8c (also called pET-3d), which was used to produce unfused HCV proteins by using the T7 expression system (66). The expression constructs and subcloning strategies are summarized in Table 1.

**HCV-specific antisera.** For production of HCV region-specific antisera, HCV polypeptides or fusion proteins expressed in *E. coli* were obtained from the insoluble fraction (5) or total cell extracts and purified by preparative sodium dodecyl sulfate (SDS)-polyacrylamide gel electrophoresis (PAGE) (5, 39). Gel slices containing the antigens were stored frozen or lyophilized and then emulsified with complete Freund's adjuvant prior to immunization of rabbits (5). Serum samples were collected after multiple booster injections with incomplete Freund's adjuvant. Reactivity and specificity of the antisera were assessed by immunoprecipitation assays with a set of radiolabeled HCV-specific cell-free translation products (data not shown; see also Results and Table 1).

Serum samples from human patients chronically infected with HCV were generously provided by Henry Hsu and Harry Greenberg (Stanford University). Coded patient designations are R, F, RJ, DS, and JHF.

**Mammalian expression constructs.** As described above, pGEM-3Zf(+) HCV(H) 17-9389F was used as the parent for mammalian expression constructs. Plasmid expression vectors were derivatives of pTM3 (provided by B. Moss) (48). pTM3 contains a unique *Nco*I site and a polycloning region immediately 3' to the T7 promoter and the encephalomyocarditis virus (EMC) internal ribosome entry site (IRES) (16). This system yields high levels of protein expression by using vTF7-3, a vaccinia virus recombinant which expresses the T7 DNA-dependent RNA polymerase (20). In addition, pTM3 contains flanking vaccinia virus DNA and a dominant selectable marker which readily allow rescue of the corresponding vaccinia virus recombinants (see below).

TABLE 2. HCV mammalian expression constructs

| Construct[a] | HCV restriction site[b] | Vector site[c] | C-terminal residue (no.)[d] |
|---|---|---|---|
| pTM3/HCV 1-3011 | *Apa*LI (335), *Xba*I[e] | ***Nco*I**, *Spe*I | 0 |
| pBRTM/HCV 1-3011 | *Apa*LI (335), *Xba*I[e] | ***Nco*I**, *Spe*I | 0 |
| pBRTM/HCV 827-3011 | *Nco*I[g], *Xba*I[e] | *Nco*I, *Spe*I | 0 |
| pBRTM/HCV 1-2940[f] | ***Apa*LI** (335), ***Nde*I** (9158) | ***Nco*I**, ***Pac*I** | 6 |
| pBRTM/HCV 1-2813 | ***Apa*LI** (335), *Nru*I (8776) | ***Nco*I**, *Stu*I | 0 |
| pBRTM/HCV 1-2508 | ***Apa*LI** (335), ***Hin*dIII** (7861) | ***Nco*I**, ***Pac*I** | 0 |
| pBRTM/HCV 1-2398 | ***Apa*LI** (335), ***Bam*HI** (7529) | ***Nco*I**, *Stu*I | 0 |
| pBRTM/HCV 1-2205 | ***Apa*LI** (335), *Pvu*II (6954) | ***Nco*I**, ***Pac*I** | 0 |
| pBRTM/HCV 1-2101 | ***Apa*LI** (335), *Sna*BI (6642) | ***Nco*I**, ***Pac*I** | 0 |
| pBRTM/HCV 1-2051 | ***Apa*LI** (335), ***Sse*8387I** (6493) | ***Nco*I**, ***Pac*I** | 0 |
| pBRTM/HCV 1-1957 | ***Apa*LI** (335), ***Bsu*36I** (6209) | ***Nco*I**, *Stu*I | 1 |
| pBRTM/HCV 1-1864 | ***Apa*LI** (335), ***Bsm*I** (5934) | ***Nco*I**, *Stu*I | 3 |
| pBRTM/HCV 1-1773 | ***Apa*LI** (335), *Ssp*I (5659) | ***Nco*I**, *Stu*I | 1 |
| pBRTM/HCV 1-1692 | ***Apa*LI** (335), *Nae*I (5414) | ***Nco*I**, *Stu*I | 3 |
| pBRTM/HCV 1-1676[f] | ***Apa*LI** (335), *Hin*cII (5366) | ***Nco*I**, *Stu*I | 1 |
| pBRTM/HCV 1-1546 | ***Apa*LI** (335), *Sma*I (4976) | ***Nco*I**, *Stu*I | 3 |
| pTM3/HCV 1-1488 | ***Apa*LI** (335), *Bam*HI[g,h] | ***Nco*I**, *Pst*I | 0 |

[a] The numbers refer to the portion of the HCV polyprotein encoded by each construct. Flanking residues present in the polyproteins are not included. For all of the constructs except pBRTM/HCV 827-3011, three additional N-terminal residues (Met-Cys-Thr) are predicted to be present prior to the Met residue initiating the HCV polyprotein (see Materials and Methods).

[b] Restriction sites in the HCV cDNA used for the plasmid constructs. Nucleotide numbers in parentheses refer to the positions of these sites in the full-length HCV H-strain sequence (14), assuming that the 5′ noncoding region of HCV H and HCV type 1 (8, 28) are the same length. Restriction sites shown in bold type indicate that protruding ends were treated with either the Klenow fragment of DNA polymerase I or T4 DNA polymerase prior to ligation to produce blunt ends. All constructs contained the expected sequences at the ligation junctions except as noted above.

[c] Restriction sites in plasmid vectors used for cloning. Sites shown in bold type indicate that protruding ends were treated with either the Klenow fragment of DNA polymerase I or T4 DNA polymerase prior to ligation to produce blunt ends.

[d] Number of predicted non-HCV C-terminal residues prior to the first termination codon.

[e] This *Xba*I site is in the pGEM-3Zf(+) HCV(H) 17-8958 polylinker.

[f] In both pBRTM/HCV 1-2940 and pBRTM/HCV 1-1676, two nucleotides were found to be deleted at the 3′ ligation junction during subcloning.

[g] For constructs beginning with residue 827 or ending with residue 1488, the restriction sites used for cloning, *Nco*I and *Bam*HI, respectively, were derived by PCR.

[h] In the case of pTM3/HCV 1-1488 the HCV cDNA fragment containing the PCR-engineered termination codon and *Bam*HI site was obtained from pH2JC1/HCV N3.8C2 as an *Eco*O47III(2848)-*Nsi*I fragment which was subcloned into pTM3/HCV 1-966 (*Eco*O47III-*Pst*I) (27).

pTM3/HCV 1-3011 was constructed by several subcloning steps and contains the entire HCV ORF and short 5′ and 3′ flanking sequences (cDNA from nucleotide 336 to 9389). The HCV cDNA sequence is located immediately 3′ to the *Nco*I site of pTM3 which had been filled in with the Klenow fragment of *E. coli* DNA polymerase I. The 5′ DNA sequence (5′-. . . CCATGTGCACCATGA. . . -3′) contains the ATG corresponding to the preferred translation initiation site of the EMC IRES followed, in frame, by the ATG which initiates the long HCV ORF. This results in the addition of three non-HCV amino acid residues to the N terminus of the predicted translation product. The 3′ flanking sequence is 5′-. . . TGAACGGGGAGCTAGGGGATCCTCTAGT. . . -3′, where the termination codon of the HCV ORF is shown in boldface and the underlined nucleotides correspond to the remnants of the pTM3 *Spe*I restriction site inactivated during subcloning. pTM3/HCV 1-1488 contains an amber termination codon, engineered by PCR, following residue 1488 of the HCV ORF.

Since near-full-length HCV clones in pTM3 were found to be difficult to propagate, a plasmid derivative with a lower copy number and tetracycline resistance was constructed by using pBR322 (designated the pBRTM series). The *Xba*I-*Pvu*I fragment of pTM3/HCV 336-9389F, which contains the HCV cDNA insert as well as the flanking vaccinia virus sequences, was inserted into pBR322 (2) which had been digested with *Dra*I and *Eco*RI. Prior to ligation, both DNA fragments were treated with T4 DNA polymerase in the presence of deoxynucleoside triphosphates. In the parental plasmid, pBRTM/HCV 1-3011, the HCV cDNA coding sense is oriented in the same direction as *tet* transcription. The subcloning strategies for other pBRTM expression constructs used in these studies are summarized in Table 2.

Two expression constructs, pTM3/HCV 1-1488 and pBRTM/HCV 827-3011, were used to construct vaccinia virus recombinants. The corresponding vaccinia virus-HCV recombinants, vHCV 1-1488 and vHCV 837-3011, were generated by marker rescue on CV-1 cells (42) and identified by using the *gpt* selection method (17). Recombinant viruses were plaque purified three times under selective conditions prior to growth of large-scale stocks.

**Vaccinia virus transient-expression assays.** For expression assays utilizing vaccinia virus-HCV recombinants, the indicated cell types were infected with vTF7-3 alone or in combination with vHCV1-1488 or vHCV 827-3011 by using a multiplicity of infection of 5 PFU per cell of each recombinant (as determined on BSC-40 monolayers). After 30 min at room temperature, the inoculum was removed and replaced with MEM containing 2% FBS. At 2 h postinfection, monolayers were washed once with prewarmed MEM lacking methionine and labeled by incubation for 4 h at 37°C with MEM containing 1/40 the normal concentration of methionine, 2% FBS, and 50 μCi of $^{35}$S-translabel (ICN) per ml.

Expression assays of transfected plasmid constructs utilized subconfluent monolayers of BHK-21 cells in 35-mm dishes (approximately $10^6$ cells) which had been previously infected with vTF7-3 (5 to 10 PFU per cell) in 0.2 ml MEM for 30 min at 37°C. After removal of the inoculum, cells were transfected at 37°C by using a mixture consisting of 1 μg of plasmid DNA and 12.5 μg of Transfectam (Promega) in 0.5 ml of MEM. After 2.5 h, the transfection mixture was removed and the cells were incubated for 4 h at 37°C in 0.5

ml of MEM containing 1/40 the normal concentration of methionine, 2% FBS, and 40 μCi of $^{35}$S-translabel (ICN) per ml.

**Cell lysis, immunoprecipitation, and protein analyses.** Labeled monolayers were washed with phosphate-buffered saline, lysed with a solution of 0.5% SDS containing 20 μg of phenylmethylsulfonyl fluoride per ml (~0.3 ml per $10^6$ cells), and sheared by repeated passage through a 26-gauge needle. If the lysates were not used immediately, aliquots were stored frozen at −70°C. Before use, samples were heated at 70°C for 15 min, diluted into the immunoprecipitation buffer containing Triton X-100 and carrier bovine serum albumin (57), and clarified by centrifugation at 16,000 × *g* for 15 min. Portions of each lysate were incubated with the indicated HCV region-specific antisera (usually 5 μl), and the immune complexes were collected by using *Staphylococcus aureus* Cowan strain I (Calbiochem) as described previously (57). Immunoprecipitates were solubilized and analyzed by SDS-PAGE (39, 63). After treatment for fluorography by using either diphenyloxazole (40) or En$^3$Hance (Du Pont), gels were dried and exposed at −70°C by using prefogged (40) X-ray film (Kodak). The apparent molecular weights of HCV-specific antigens were estimated by comparison with $^{14}$C-methylated marker proteins (Amersham).

## RESULTS

**Identification of HCV polyprotein processing products.** To examine the synthesis and processing of HCV proteins in mammalian cell cultures, a series of expression plasmids were assembled from HCV H strain cDNA clones (Fig. 1C). These constructs were designed for expression of HCV polyproteins by using the vaccinia virus-T7 system, either by using a plasmid transfection protocol or by rescue of vaccinia virus-HCV recombinants (vHCV). Uncapped mRNA transcripts were made by the T7 DNA-dependent RNA polymerase encoded by vaccinia virus recombinant vTF7-3 and contained the EMC 5′ IRES in order to achieve efficient cap-independent translation of HCV coding regions. To identify HCV-specific polyproteins and cleavage products, subregions of the HCV polyprotein were expressed in *E. coli*, purified, and used to produce a panel of polyclonal rabbit antisera suitable for immunoprecipitation of SDS-denatured HCV antigens (Table 1; Fig. 1B).

Two vaccinia virus recombinants encoding overlapping polyproteins were used to produce the [$^{35}$S]Met-labeled HCV-specific products shown in Fig. 2. vHCV1-1488 is predicted to express an HCV polyprotein which initiates with 3 extra N-terminal residues followed by the first 1,488 residues of the HCV ORF (Fig. 1). The second recombinant, vHCV827-3011, begins at Met-827 in the putative NS2 region and extends to the end of the HCV ORF. These two recombinants were used since preliminary experiments indicated that C-terminal HCV polypeptides were underproduced in cells expressing the entire HCV polyprotein (either infected with vHCV1-3011 or transfected with pBRTM/HCV 1-3011). However, products of identical immunoreactivity and size identical to those shown in Fig. 2 are observed in expression studies with constructs expressing the full-length polyprotein (26).

Figure 2A shows the products immunoprecipitated from SDS-denatured extracts from vHCV-infected BHK-21 cells. Antiserum WU120, directed against a fusion protein containing HCV polyprotein residues 1 to 142, immunoprecipitated a protein of 21 kDa which is thought to represent the HCV capsid protein. Antiserum WU122, which was directed against the putative HCV E1 envelope protein sequences, weakly immunoprecipitated a diffuse 31-kDa protein. Surprisingly, this polypeptide was also immunoprecipitated by antisera WU105 and WU107, which are directed against the

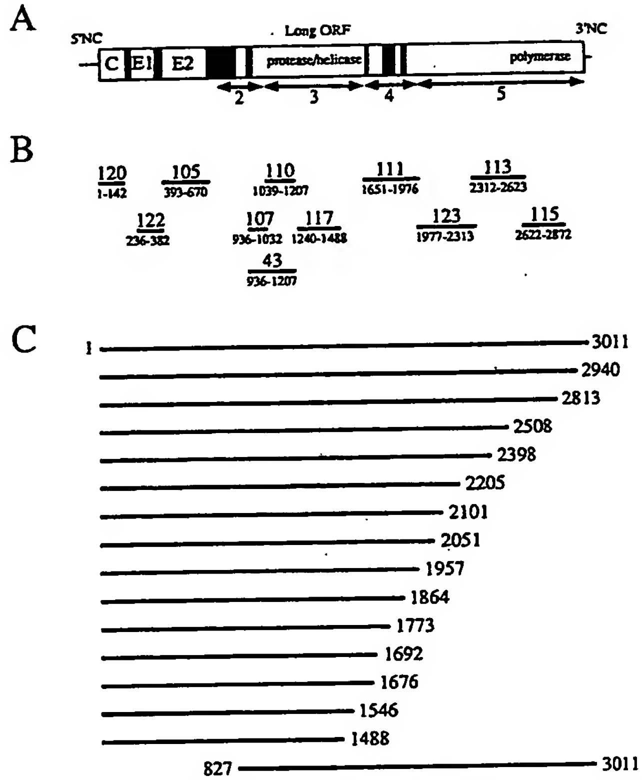
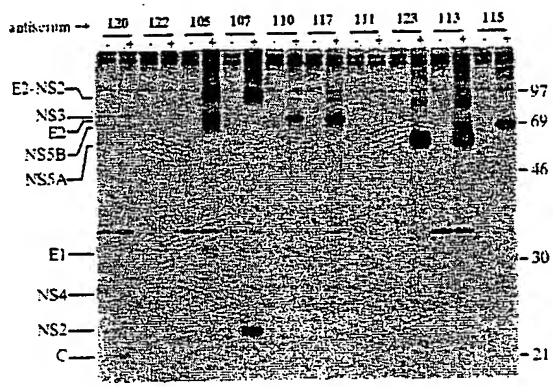



FIG. 1. HCV genome structure, region-specific antisera, and expression constructs. (A) Diagram of the HCV H strain genome RNA is shown with the 5′ and 3′ noncoding regions (NC) indicated by lines and the long ORF denoted as a box. It is not known if the H strain genome RNA contains a 5′-terminal cap structure or a 3′-terminal poly(A) (28) or poly(U) (36, 52, 67) tract. The locations of the putative structural proteins (30), the basic capsid protein (C) and two envelope glycoproteins, E1 and E2, are shown. Regions of the polyprotein containing predominantly uncharged amino acids are indicated as black bars. In this report, the nomenclature used to describe the remaining regions of the HCV polyprotein is based on that of the flaviviruses (for a review, see reference 4) and assumes similar functional organization (8, 47, 67). It appears that HCV may not encode a protein analogous to the secreted nonstructural NS1 glycoprotein of flaviviruses (64). Following E2, the HCV polyprotein contains a hydrophobic portion, like the NS2 region of flaviviruses, which precedes a putative serine proteinase domain (1) and NTPase and helicase motifs (23) which are present in the flavivirus NS3 protein (58, 74). Following the NS3 region is another hydrophobic region called NS4. The remaining portion of the ORF is referred to as the NS5 region, and the C-terminal part of this coding sequence contains the Gly-Asp-Asp motif characteristic of RNA-dependent-RNA polymerases (53). (B) The portions of the HCV polyprotein used as immunogens for production of polyclonal rabbit antisera are indicated as black lines. Above each line is the designation for each antiserum as used in this report; below each line is the region of the polyprotein present in each expression construct (numbered from the first Met residue in the long HCV ORF). See Table 1 and Materials and Methods for further details. (C) Summary of the HCV polyprotein expression constructs used in this study (Table 2). Polyprotein sequences present in each construct are indicated by black lines which are drawn to scale and oriented with respect to the diagram of the HCV genome shown in panel A.

A




B

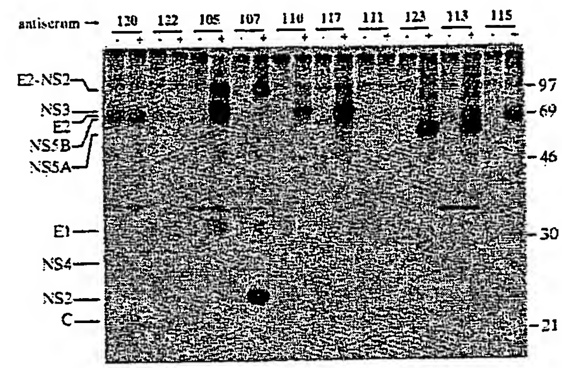



FIG. 2. Identification of HCV polyprotein cleavage products. BHK-21 monolayers were infected with vTF7-3 alone (−) or coinfected with vTF7-3 and either vHCV 1-1488 or vHCV 827-3011 (+) and labeled with $^{35}$S-translabel as described in Materials and Methods. Cell lysates were prepared and immunoprecipitated by using the indicated HCV region-specific antisera, as described in Table 1 and Fig. 1. As discussed in Results, lysates from cells coinfected with vHCV 1-1488 were used for immunoprecipitations with region-specific antisera WU120, WU122, WU105, and WU107. Lysates from cells coinfected with vHCV 827-3011 were used for immunoprecipitations with region-specific antisera WU110, WU117, WU111, WU123, WU113, and WU115. Samples were separated on SDS–14% polyacrylamide gels. (A) Lysates prepared from BHK-21 cells; (B) lysates prepared from HepG2 A16 cells. HCV-specific proteins are identified at the left of each panel with the sizes of protein molecular weight markers indicated at the right.

E2 and NS2 regions of the polyprotein, respectively (see Discussion). WU105 antiserum also immunoprecipitated diffuse products of 88 and 70 kDa (the smaller product being the putative E2 glycoprotein). The 88-kDa product was also immunoprecipitated by WU107 antiserum, which is directed against the NS2 region, and is therefore called E2-NS2. A predominant 23-kDa product also reacted with the NS2 region-specific antiserum and is referred to as NS2. Antiserum directed against either the putative serine proteinase domain (WU110) or the helicase-NTPase domain (WU117) specifically immunoprecipitated a 70-kDa protein, NS3, which is nearly identical in size to the homologous flavivirus protein. The HCV NS4 region-specific antiserum (WU111) reacted weakly with a 27-kDa product. (The 22-kDa HCV-specific product which is also present in this lane [as well as some of the other lanes] is an N terminally truncated form of the NS2 protein produced by vHCV 827-3011 which often precipitates nonspecifically.) Antiserum directed against the N-terminal portion of the NS5 region (WU123) precipitated a predominant species of 58 kDa, with additional minor slower mobility forms (up to 68 kDa, but difficult to see in the exposures shown in Fig. 2), which are collectively called NS5A. Antiserum to the middle portion of the NS5 region (WU113) reacted with a 68-kDa polypeptide in addition to NS5A. This species, called NS5B, was also immunoprecipitated by antiserum directed against the C-terminal portion of the NS5 region (WU115). Besides these major species, larger polypeptides, consistent with uncleaved polyproteins, were identified with several of the antisera (in particular, see WU110, WU117, WU123, WU113, and WU115). We are currently studying the kinetics of HCV polyprotein processing (41) and the identification of these possible processing intermediates will be discussed in detail elsewhere.

Since hepatocytes are believed to be permissive for HCV infection and replication (49), we also examined the HCV-specific proteins expressed in the human hepatoma cell line, HepG2 A16. Dramatic host-specific differences in processing were not observed, and results essentially identical to those discussed above for BHK-21 cells were obtained (Fig. 2B). Similar patterns of processed products were also found in CV-1 (monkey kidney) and CHO (hamster ovary) cells (data not shown).

**N-linked glycosylation of HCV polypeptides.** To determine whether any of the HCV-specific proteins expressed by using this system contained asparagine-linked carbohydrate, lysates of $^{35}$S-labeled BHK-21 cells were immunoprecipitated with each region-specific antiserum and digested with endoglycosidase F, which removes both high-mannose and complex glycans (15). The only HCV-specific polypeptides converted to faster migrating forms by endoglycosidase F digestion were E1, E2, and E2-NS2, suggesting that they contain N-linked glycans (Fig. 3). E1 was converted from a 31-kDa species to a 21-kDa deglycosylated form. The pattern of E2-specific products was more complex, and at least three E2-specific endoglycosidase F digestion products were observed. The largest product (62 kDa) was also present in the sample immunoprecipitated with NS2 region-specific antiserum (WU107), and hence probably represents deglycosylated E2-NS2. The two other discrete species, of 41 and 36 kDa, presumably represent deglycosylated forms of E2. Whether these multiple forms reflect different E2 polypeptide backbones or result from other posttranslational modifications is unknown (see Discussion). As mentioned above, although SDS-denatured lysates are heated before immunoprecipitation, E1 coprecipitates with the E2 and NS2 region-specific sera, and deglycosylated forms of E1 can be observed in these samples. These results are consistent with previous expression studies which have indicated that HCV E1 and E2 produced in cell-free translation systems (30) or mammalian (37, 45, 64) or insect cells (45) are both heavily modified by N-linked glycosylation. The predicted sequences of the HCV H strain E1 (polyprotein residues 192 to 383 [30]) and E2 (polyprotein residues 384 to ~750 [30]) proteins contain 5 and 9 potential acceptor sites, respectively, and our data indicate that the majority of these sites are utilized in mammalian cells (assuming 2 to 3 kDa per oligosaccharide unit).

**HCV antigens recognized by human sera.** Although the majority of the HCV polyprotein was represented in the antigens used for production of region-specific sera, immunogens from some regions have not be obtained (note the NS2 region in Fig. 1), and the immunodominant epitopes

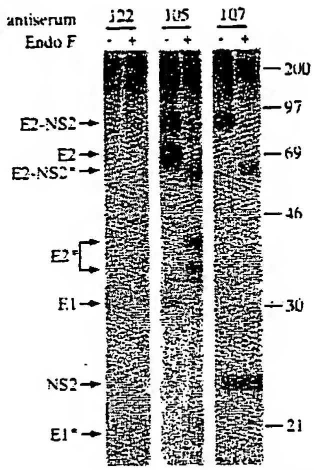



FIG. 3. Endoglycosidase F digestion of HCV glycoproteins. Cell monolayers were coinfected with vTF7-3 and vHCV 1-1488 and labeled with $^{35}$S-translabel as described in Materials and Methods. Equivalent portions of the cell lysate were immunoprecipitated with WU122 (E1 specific), WU105 (E2 specific), or WU107 (NS2 specific). Immunoprecipitates were resuspended and incubated overnight in either the absence (−) or presence (+) of endoglycosidase F. Digestions were conducted essentially as previously described (44 [and references therein]). Samples were separated on SDS–14% polyacrylamide gels. HCV-specific proteins are identified at the left, and the sizes of protein molecular weight markers are indicated at the right. The positions of endoglycosidase F-digested forms (indicated by asterisks) are also indicated. Although not shown, parallel samples were analyzed from vTF7-3-infected monolayers to unambiguously allow identification of the HCV-specific products (see also Fig. 2).

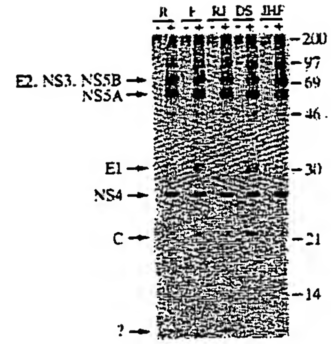



FIG. 4. Immunoprecipitation of HCV antigens with human antisera. Lysates from cells infected with vTF7-3 alone (−) or a mixture of lysates from cells coinfected with vTF7-3 and either vHCV 1-1488 or vHCV 827-3011 (+) were used for immunoprecipitation with serum from five different HCV seropositive patients (denoted R, F, RJ, DS, and JHF. Samples were separated on SDS–14% polyacrylamide gels. HCV-specific proteins are identified at the left, and the sizes of protein molecular weight markers are indicated at the right.

recognized by our panel of sera have not been defined. Hence, some HCV-encoded polypeptides may have been missed in our analyses. In the hope of identifying additional HCV-specific cleavage products, we examined the reactivity of serum samples from five HCV-infected patients. Radiolabeled SDS-denatured lysates were prepared from BHK-21 cells infected with vTF7-3 alone or coinfected with either vHCV 1-1488 or vHCV 827-3011. Coinfected lysates which had been pooled were used for the immunoprecipitation analyses shown in Fig. 4. All patient sera showed strong reactivity with NS5A and the 27-kDa NS4 product and various degrees of reactivity with E1 and C. None of the sera showed detectable reactivity with NS2 or its truncated form. A strong band migrating at ~70 kDa was immunoprecipitated by all patient sera, but since E2, NS3, and NS5B all migrate in this size range it is difficult to interpret these results. In addition to these species, a small HCV-specific polypeptide of ~8 kDa, which had not been previously identified, was immunoprecipitated by serum from patients R, RJ, DS (weak reaction), and JHF. A similarly sized species was also observed in longer exposures of immunoprecipitations of vHCV 827-3011-infected lysates with the NS4 region-specific antiserum (WU111), suggesting that this product is derived from additional processing of the NS4 region (see below).

**Fine mapping the positions of HCV nonstructural protein by deletion analyses.** The reactivity of the HCV proteins with region-specific antisera can be used to roughly map their locations in the HCV polyprotein. The sizes and immunoreactivities of C, E1, and E2 from our studies are consistent with previous results from cell-free translation studies which defined cleavages after residues 191 and 383, dependent on microsomal membranes, to produce the N termini of E1 and E2, respectively (30). For the NS2 region, a truncated form of NS2 is produced in lysates infected with vHCV 827-3011 (data not shown). This truncated product is about 1.5 kDa smaller than the 23-kDa NS2 protein, which suggests that the N terminus of NS2 is produced by cleavage in the vicinity of residues 805 to 815. In the case of NS3, alignment with the homologous flavivirus NS3 proteins predicts a cleavage in the vicinity of residues 1020 to 1030, which is consistent with the observed reactivity of HCV NS3 with antiserum directed against the fusion protein encompassing residues 1039 to 1207 but not the antiserum directed against the NS2 region (residues 936 to 1032). The locations of the remaining cleavage sites generating the putative HCV nonstructural proteins are less well defined. As mentioned above, NS4 region antiserum (WU111) recognized two proteins of 27 and 8 kDa, but their order in the polyprotein cannot be established from these experiments. Finally, while the order of NS5A and NS5B in the polyprotein is clear from the experiments presented in Fig. 2, the location of the 4-5A and 5A-5B cleavage sites can only be roughly localized on the basis of the apparent sizes and immunoreactivity of these products.

A more precise map of the locations of the putative HCV nonstructural proteins was obtained by examining the cleavage products from a series of polyproteins with C-terminal truncations (diagrammed in Fig. 1). It should be noted that all of these constructs contained the putative HCV serine proteinase domain (approximately residues 1020 to 1207), which has been shown to be required for downstream proteolytic processing (25). Hence, if a C-terminal deletion does not affect processing at the normal cleavage sites, the sizes of truncated products should allow rough mapping of the C-terminal boundaries of HCV proteins. Polyproteins terminating at residues 2940 or 2813 produced truncated NS5B-specific products of 63 or 43 kDa, respectively, compared with the 68-kDa NS5B species produced by the full-length polyprotein (Fig. 5A). Normal NS3, NS4 region products (data not shown), and NS5A (Fig. 5B) were produced by these constructs, which suggests that NS5B is the C-terminal HCV nonstructural protein with its C terminus located at or very near the end of the HCV ORF. Polypro-

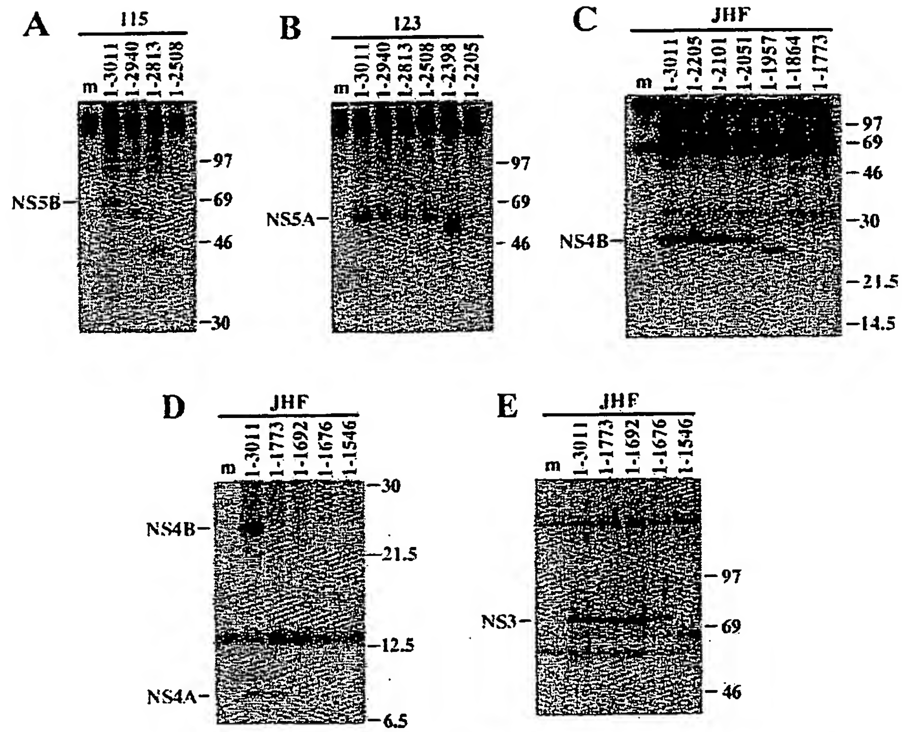



FIG. 5. C-terminal boundaries of the HCV nonstructural proteins. A series of constructs encoding progressive C-terminal deletions (Table 2 and Fig. 1) were used to map the C-terminal boundaries of NS5B (A), NS5A (B), NS4B (C), NS4A (D), and NS3 (E). BHK-21 cells previously infected with vTF7-3 were transfected with the indicated plasmid DNAs or mock transfected (m) and labeled with $^{35}$S-translabel as described in Materials and Methods. Cell lysates were prepared, and HCV-specific antigens were immunoprecipitated by using rabbit antiserum WU115 (A), WU123 (B), or human serum from patient JHF (C, D, and E). Immunoprecipitated proteins were separated by electrophoresis on 10% (A and B), 14% (C), or 8% (E) polyacrylamide-SDS gels or a 14% polyacrylamide tricine gel (D). In the case of the NS5A-specific (WU123) immunoprecipitation products (B), it should be noted that a nonspecific protein comigrates with the predominant form of the NS5A protein. HCV-specific proteins are identified at the left of each panel, and the sizes of protein molecular weight markers are indicated at the right.

teins terminating at residues 2398, 2205, 2101, or 2051 produced NS3 and NS4 region species identical to the full-length polyprotein (data not shown; also see Fig. 5C). However, the construct terminating at residue 2398 produced an NS5A-specific species of 53 kDa, truncated by about 5 kDa (Fig. 5B). This suggests that the 5A-5B cleavage site lies between residues 2398 and 2508 (near residue 2440, on the basis of the size of the truncated NS5A product). Polyproteins terminating at residues 1957 and 1864 produced truncated NS4 forms of 25 and 14 kDa (Fig. 5C), respectively, mapping the cleavage site producing the C terminus of the 27-kDa NS4 region species to between residues 1957 and 2051 (near residue 1975). This protein is subsequently referred to as NS4B. Analyses to define the C-terminal boundaries of the 8-kDa NS4 region species and NS3 are shown in Fig. 5D and E. The polyprotein terminating at residue 1773 produced the 8-kDa NS4 region species, but this product disappeared after truncation to residue 1692 (Fig. 5D). This species is subsequently referred to as NS4A, and the data suggest that the 4A-4B cleavage site lies between residues 1692 and 1773. The polyprotein terminating at residue 1692 produced normal NS3; however, the construct truncated to residue 1676 produced a slightly larger (~1-kDa) form, suggesting that 3-4A cleavage had been blocked (Fig. 5E). The 1-1546 polyprotein produced a 63-kDa form of the NS3 protein, truncated by about 7 to 8 kDa. These data are consistent with NS3-NS4A cleavage between residues 1546 and 1676, probably near residue 1665.

Although N- and C-terminal sequence analyses will be needed to define the precise boundaries of the HCV polyprotein cleavage products, these results establish a preliminary map of the HCV H strain-encoded polypeptides and cleavage sites. The HCV polyprotein organization defined by these expression studies is C(p21)-E1(gp31)-E2(gp70)-?-NS2(p23)-NS3(p70)-NS4A(p8)-NS4B(p27)-NS5A(p58)-NS5B(p68). These results are summarized in the diagram shown in Fig. 6, in which the sizes and locations of the cleavage products are drawn to scale on the basis of the data presented in this paper and elsewhere (25, 30).

## DISCUSSION

The vaccinia virus transient-expression system has been used for numerous studies examining processing of RNA virus polyproteins. In general, the results from these studies mimic the authentic processing reactions observed in virus-infected cells. However, since an efficient cell culture replication system is lacking for HCV, such a comparison is not yet possible. With this caveat in mind, several points have emerged from our studies.

Consistent with previous studies (6, 29, 37, 38, 45, 64), expression of the full-length HCV polyprotein or truncated derivatives containing the putative structural region led to the production of a 21-kDa N-terminal product believed to represent the HCV capsid protein and two glycoproteins which were heavily modified by N-linked glycosylation,

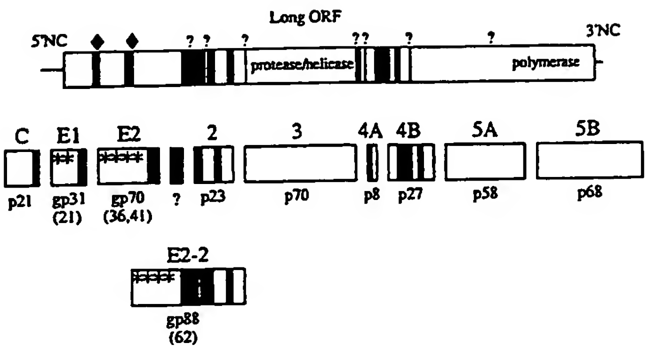



FIG. 6. Summary of HCV polyprotein processing products. HCV cleavage products as identified in this vaccinia virus transient-expression study are indicated below a diagram of the HCV polyprotein (labeled as in Fig. 1). Putative cleavage sites for host signalase identified by cell-free translation studies (30) are indicated by filled diamonds. In the top diagram, sites of polyprotein cleavage mediated by unknown proteinases are indicated (?). The nomenclature used for HCV polypeptides follows that of the flaviviruses (55, 56). The observed sizes for HCV proteins (p) and glycoproteins (gp) are indicated. For the glycoproteins (E1, E2, and E2-NS2), the sizes of the endoglycosidase F-resistant forms are given in parentheses. Although not identified in this study, the NS2 region may encode an additional product(s) (?). The apparent molecular mass of E2-NS2 is 88 kDa, as measured by SDS-12% PAGE, but it migrates as an 82-kDa species on 8% polyacrylamide gels. Asterisks denote proteins with N-linked glycans but do not necessarily indicate the position or number of sites utilized. See the text for further discussion.

designated E1 (31 kDa) and E2 (70 kDa). In addition to these products, an 88-kDa glycoprotein reacting with both E2 and NS2 region-specific antisera was identified. Preliminary studies indicate that this E2-NS2 protein may represent a precursor to E2 (24). However, several observations suggest that processing in the E2-NS2 region is complex. Endoglycosidase F digestion of E2, which migrates as a broad band on our SDS gels, produces at least two species of 36 and 41 kDa (obvious heterogeneity is also apparent in the 41-kDa product). The relationship of these two species is not yet clear. In addition, it is difficult to propose a simple processing model for this region on the basis of the apparent sizes of the E2 species, NS2, and E2-NS2. The NS2 protein is 23 kDa, with its C terminus predicted in the vicinity of residues 1020 to 1030. Given that cell-free translation studies have defined the putative signalase cleavage site generating the E2 N terminus after residue 383 (30), this predicts that the polypeptide backbone of an E2-NS2 precursor should be ~71 kDa. However, the endoglycosidase F digestion product derived from the 88-kDa E2-NS2 glycoprotein is only 62 kDa. If the N-terminal residue of E2-NS2 is 384, then a polypeptide with a predicted size of 62 kDa would barely overlap with the NS2 region used to produce our NS2 region-specific antiserum. Although these products may simply exhibit aberrant migration on SDS-polyacrylamide gels, it is also possible that alternative proteolytic processing or other posttranslational modifications are occurring in the E2-NS2 region, leading to the production of multiple forms of E2 with possibly distinct biological functions in HCV replication. Alternatively, the two endoglycosidase F-resistant forms of E2 could reflect a delayed cleavage in the maturation of HCV E2, similar to those observed for the spike glycoproteins of many enveloped viruses (65). Interestingly, pulse-chase studies with a CHO cell line expressing the HCV structural region indicate that the 70-kDa form of E2 is chased to a 68-kDa species (64), although the nature of this modification has not been defined. Finally, from the observed versus the predicted sizes of the various E2 and NS2 species, it is also possible that one or more polypeptides from the NS2 region have gone undetected in our studies (labeled ? in Fig. 6). In this regard, a 10-kDa HCV-specific product appears to be recognized by several human sera when a nonionic detergent, rather than SDS, is used for preparation of cell lysates (data not shown). Additional immunological reagents, kinetic analyses, and N-terminal sequence data will be necessary to further clarify processing in this region.

An intriguing observation was that E1 was coprecipitated by both E2 and NS2 region-specific antisera (Fig. 2 and 4), even though samples were denatured by heating in SDS prior to immunoprecipitation. Prior reduction of samples with dithiothreitol dramatically reduced the amount of E1 associated with E2 or E2-NS2, suggesting that these proteins might be linked, either directly or via other proteins, by disulfide bonds (24). This point is of particular interest, since previous studies with hog cholera virus, a pestivirus, have shown that the three virion glycoproteins are present as disulfide-linked homodimers (gp44/48 and gp55) and heterodimers (gp35-gp55) in purified virions (69), in infected cells (70), and when expressed via a vaccinia virus recombinant (59). This is in contrast to the flavivirus West Nile virus (WN), in which the envelope protein precursors, prM and E, are associated as a stable heterodimer in nonionic detergent, but not by disulfide bridges (73). Although further studies are needed to clarify the situation with HCV, preliminary analyses with the vaccinia virus expression system indicate that some but not all of the putative virion envelope proteins are associated as discrete disulfide-linked oligomers which are formed soon after synthesis (24).

The HCV-specific proteins produced by processing of the remainder of the polyprotein are remarkably similar to those of the pestiviruses and flaviviruses (for reviews, see references 4, 10, and 58). The 70-kDa HCV NS3 protein is nearly identical in size to the homologous NS3 protein of flaviviruses (p80 in the case of some pestiviruses). For flaviviruses (58) and pestiviruses (75), the N-terminal one-third of this protein has been shown to function as a serine proteinase mediating several cleavages in the viral polyproteins, and similar results have now been obtained for HCV (25). The remainder of the protein contains motifs characteristic of NTPases and helicases (23), and for WN virus this domain has been shown to possess NTPase activity (74). For strains of the pestivirus bovine viral diarrhea virus (BVDV), a fascinating correlation has been made between production of p80 and virus-induced cytopathic effect and fatal mucosal disease (46 [and citations therein]). In noncytopathic strains, the cleavage producing the N terminus of p80 does not occur, resulting in the production of p125 (p54-p80 polyprotein). In cytopathic strains isolated from animals with mucosal disease, which sometimes occurs after congenital transmission of BVDV, insertion of host sequences and/or duplications in the BVDV genome RNA allow production of both p125 and p80 (46). Thus far, it appears that processing at the equivalent cleavage site (2-3) occurs efficiently in flavivirus-infected cells (58) and for the HCV H strain (at least as assayed by transient expression using vaccinia virus). Whether the efficiency of the 2-3 cleavage in these viruses will also correlate with the severity of cytopathic effects and pathogenesis remains to be determined. Given the growing number of divergent HCV isolates (68; see

reference 33 for a review) and the high mutation rate and evolution of this virus during chronic infection of primate hosts (43, 50, 51, 71), it will be of interest to see if strain-specific differences in processing which correlate with clinical disease can be found.

Immediately C terminal to the NS3 protein, the relatively hydrophobic NS4 region is processed to yield two proteins, NS4A (8 kDa) and NS4B (27 kDa). Similar processing events occur in this region of the flavivirus and pestivirus polyproteins, but the function of these small proteins in virus replication is unknown. For most flaviviruses, NS4A has been difficult to identify because of a lack of immune reagents and possible instability of the protein, but this region of the polyprotein encompasses about ~16 kDa of protein coding sequence. In pestiviruses, the corresponding region is processed to yield the p10 protein. The following protein, NS4B or p30 for the pestiviruses (46), appears to be similar in size for all three flavivirus genera. Finally, two proteins are derived from the HCV NS5 region, NS5A (58 kDa) and the C-terminal NS5B product (66 kDa). The HCV NS5B protein is predicted to contain the RNA-dependent RNA polymerase activity on the basis of the presence of the characteristic Gly-Asp-Asp sequence (residues 2737 to 2739) and surrounding conserved motifs (53). This is similar to the case for pestiviruses, which also produce two cleavage products, called p58 and p75. In flaviviruses, the NS5 region is not further processed but remains as a single polypeptide of ~100 kDa (905 residues in the case of yellow fever virus).

The experiments reported here have given us a preliminary picture of HCV polyprotein organization and processing. However, this view is far from complete, and additional studies are needed to define polyprotein cleavage sites and the responsible proteinases and to verify that the products observed in these expression studies are similar to those produced in authentic HCV infections. Such information should prove valuable for expression and characterization of HCV-encoded enzymes as potential targets for antiviral therapy and will allow future studies to be undertaken to assess the involvement of individual HCV polypeptides in the establishment of chronic infections, virus-induced cytopathic effects (if this is the case), and evasion of immunological surveillance and to determine if these proteins play a direct role in the association of HCV with hepatocellular carcinoma.

## ACKNOWLEDGMENTS

We thank Dorota Paluszynska for expert technical assistance; Bernard Moss and F. William Studier for plasmid vectors; Henry Hsu and Harry Greenberg for human sera; Alan Schwartz for the HepG2 A16 cell line; many colleagues, especially John Majors, for helpful discussions during the course of this work; and Peter Bredenbeek, Jean Dubuisson, Mark Heise, Julie Lemm, and Margaret MacDonald for critical reading of the manuscript.

This work was supported by grants from the Pew Memorial Trusts and the Public Health Service (CA57973). C.L. is a predoctoral candidate and was supported in part by the Division of Biology and Biomedical Sciences at Washington University.

## REFERENCES

1. **Bazan, J. F., and R. J. Fletterick.** 1990. Structural and catalytic models of typsin-like viral proteases. Sem. Virol. 1:311–322.
2. **Bolivar, F., R. L. Rodriguez, P. J. Greene, M. C. Betlach, H. L. Heyneker, H. W. Boyer, J. H. Crosa, and S. Falkow.** 1977. Construction and characterization of new cloning vehicles: a multipurpose cloning system. Gene 2:95–113.
3. **Brockman, W. W., and D. Nathans.** 1974. The isolation of simian virus 40 variants with specifically altered genomes. Proc. Natl. Acad. Sci. USA 71:942–946.
4. **Chambers, T. J., C. S. Hahn, R. Galler, and C. M. Rice.** 1990. Flavivirus genome organization, expression, and replication. Annu. Rev. Microbiol. 44:649–688.
5. **Chambers, T. J., D. W. McCourt, and C. M. Rice.** 1989. Yellow fever virus proteins NS2A, NS2B, and NS4B: identification and partial N-terminal amino acid sequence analysis. Virology 169: 100–109.
6. **Chiba, J., H. Ohba, Y. Matsuura, Y. Watanabe, T. Katayama, S. Kikuchi, I. Saito, and T. Miyamura.** 1991. Serodiagnosis of hepatitis C virus (HCV) infection with an HCV core protein molecularly expressed by a recombinant baculovirus. Proc. Natl. Acad. Sci. USA 88:4641–4645.
7. **Choo, Q.-L., G. Kuo, A. Weiner, L. R. Overby, D. W. Bradley, and M. Houghton.** 1989. Isolation of a cDNA clone derived from a blood-borne non-A, non-B viral hepatitis genome. Science 244:359–362.
8. **Choo, Q.-L., K. H. Richman, J. H. Han, K. Berger, C. Lee, C. Dong, C. Gallegos, D. Coit, A. Medina-Selby, P. J. Barr, A. J. Weiner, D. W. Bradley, G. Kuo, and M. Houghton.** 1991. Genetic organization and diversity of the hepatitis C virus. Proc. Natl. Acad. Sci. USA 88:2451–2455.
9. **Choo, Q.-L., A. J. Weiner, L. R. Overby, G. Kuo, M. Houghton, and D. W. Bradley.** 1990. Hepatitis C virus: the major causative agent of viral non-A, non-B hepatitis. Br. Med. Bull. 46:423–441.
10. **Collett, M. S.** 1992. Molecular genetics of pestiviruses. Comp. Immunol. Microbiol. Infect. Dis. 15:145–154.
11. **Collett, M. S., D. K. Anderson, and E. Retzel.** 1988. Comparisons of the pestivirus bovine viral diarrhoea virus with members of the Flaviviridae. J. Gen. Virol. 69:2637–2643.
12. **Colombo, M., G. Kuo, Q.-L. Choo, M. F. Donato, E. D. Ninno, M. A. Tommasini, N. Dioguardi, and M. Houghton.** 1989. Prevalence of antibodies to hepatitis C virus in Italian patients with hepatocellular carcinoma. Lancet ii:1006–1008.
13. **Colombo, M. G., and M. G. Rumi.** 1991. HCV and hepatocellular-carcinoma: direct or indirect connection?, p. 30. *In* 3rd International Symposium on Hepatitis C Virus. Advanced Therapeutics Communications, Strasbourg, France.
14. **Daemer, R., C. Wychowski, A. Grakoui, C. M. Rice, and S. M. Feinstone.** Unpublished data.
15. **Elder, J. H., and S. Alexander.** 1982. *endo*-β-*N*-Acetylglucosaminidase F: endoglycosidase from *Flavobacterium meningosepticum* that cleaves both high-mannose and complex glycoproteins. Proc. Natl. Acad. Sci. USA 79:4540–4544.
16. **Elroy-Stein, O., T. R. Fuerst, and B. Moss.** 1989. Cap-independent translation of mRNA conferred by encephalomyocarditis virus 5′ sequence improves the performance of the vaccinia virus/bacteriophage T7 hybrid expression system. Proc. Natl. Acad. Sci. USA 86:6126–6130.
17. **Falkner, F., and B. Moss.** 1988. *Escherichia coli gpt* gene provides dominant selection for vaccinia virus open reading frame expression vectors. J. Virol. 62:1849–1854.
18. **Feinstone, S., H. J. Alter, H. P. Dienes, Y. Shimizu, H. Popper, D. Blackmore, D. Sly, W. T. London, and R. H. Purcell.** 1981. Non-A, non-B hepatitis in chimpanzees and marmosets. J. Infect. Dis. 144:588–598.
19. **Francki, R. I. B., C. M. Fauquet, D. L. Knudson, and F. Brown (ed.).** 1991. Classification and nomenclature of viruses. Fifth report of the International Committee on Taxonomy of Viruses. Arch. Virol. 1991(Suppl. 2):223.
20. **Fuerst, T. R., E. G. Niles, F. W. Studier, and B. Moss.** 1986. Eukaryotic transient-expression system based on recombinant vaccinia virus that synthesizes bacteriophage T7 RNA polymerase. Proc. Natl. Acad. Sci. USA 83:8122–8126.
21. **Gocke, D. J.** 1972. A prospective study of posttransfusion hepatitis: the role of the Australia antigen. JAMA 219:1165–1170.
22. **Gocke, D. J., H. B. Greenberg, and N. B. Kavey.** 1970. Correlation of Australia antigen with posttransfusion hepatitis. JAMA 212:877–879.
23. **Gorbalenya, A. E., E. V. Koonin, A. P. Donchenko, and V. M. Blinov.** 1989. Two related superfamilies of putative helicases

involved in replication, recombination, repair and expression of DNA and RNA genomes. Nucleic Acids Res. 17:4713–4729.

24. **Grakoui, A., C. Lin, and C. Rice.** Unpublished data.
25. **Grakoui, A., D. W. McCourt, C. Wychowski, S. M. Feinstone, and C. M. Rice.** Characterizaton of the hepatitis C virus-encoded serine proteinase: determination of proteinase-dependent polyprotein cleavage sites. Submitted for publication.
26. **Grakoui, A., B. Prágai, and C. Rice.** Unpublished data.
27. **Grakoui, A., and C. Rice.** Unpublished data.
28. **Han, J. H., V. Shyamala, K. H. Richman, M. J. Brauer, B. Irvine, M. S. Urdea, P. Tekamp-Olson, G. Kuo, Q.-L. Choo, and M. Houghton.** 1991. Characterization of the terminal regions of hepatitis C viral RNA: identification of conserved sequences in the 5′ untranslated region and poly(A) tails at the 3′ end. Proc. Natl. Acad. Sci. USA 88:1711–1715.
29. **Harada, S., Y. Watanabe, K. Takeuchi, T. Suzuki, T. Katayama, Y. Takebe, I. Saito, and T. Miyamura.** 1991. Expression of processed core protein of hepatitis C virus in mammalian cells. J. Virol. 65:3015–3021.
30. **Hijikata, M., N. Kato, Y. Ootsuyama, M. Nakagawa, and K. Shimotohno.** 1991. Gene mapping of the putative structural region of the hepatitis C virus genome by *in vitro* processing analysis. Proc. Natl. Acad. Sci. USA 88:5547–5551.
31. **Hollinger, F. B.** 1990. Non-A, non-B hepatitis viruses, p. 2239–2273. *In* B. N. Fields (ed.), Virology. Raven Press, Ltd., New York.
32. **Houghton, M., Q. Choo, and G. Kuo.** November 1988. European patent application number 88310922.5. Publication number 0318216.
33. **Houghton, M., A. Weiner, J. Han, G. Kuo, and Q.-L. Choo.** 1991. Molecular biology of the hepatitis C viruses: implications for diagnosis, development and control of viral disease. Hepatology 14:381–388.
34. **Hruby, D. E., L. A. Guarino, and J. R. Kates.** 1979. Vaccinia virus replication. I. Requirement for the host-cell nucleus. J. Virol. 29:705–715.
35. **Inchauspe, G., S. Zebedee, D.-H. Lee, M. Sugitani, M. Nasoff, and A. M. Prince.** 1991. Genomic structure of the human prototype strain H of hepatitis C virus: comparison with the American and Japanese isolates. Proc. Natl. Acad. Sci. USA 88:10292–10296.
36. **Kato, N., M. Hijikata, Y. Ootsuyama, M. Nakagawa, S. Ohkoshi, T. Sugimura, and K. Shimotohno.** 1990. Molecular cloning of the human hepatitis C virus genome from Japanese patients with non-A, non-B hepatitis. Proc. Natl. Acad. Sci. USA 87:9524–9528.
37. **Kohara, M., K. Tsukiyama-Kohara, N. Maki, K. Asano, K. Yamaguchi, K. Miki, S. Tanaka, N. Hattori, Y. Matsuura, I. Saito, T. Miyamura, and A. Nomoto.** 1992. Expression and characterization of glycoprotein gp35 of hepatitis C virus using recombinant vaccinia virus. J. Gen. Virol. 73:2313–2318.
38. **Kumar, U., D. Cheng, H. Thomas, and J. Monjardino.** 1992. Cloning and sequencing of the structural region and expression of putative core gene of hepatitis C virus from a British case of chronic sporadic hepatitis. J. Gen. Virol. 73:1521–1525.
39. **Laemmli, U. K.** 1970. Cleavage of structural proteins during the assembly of the head of bacteriophage T4. Nature (London) 227:680–685.
40. **Laskey, R. A., and A. D. Mills.** 1975. Quantitative film detection of $^3H$ and $^{14}C$ in polyacrylamide gels by fluorography. Eur. J. Biochem. 56:335–341.
41. **Lin, C., A. Grakoui, and C. Rice.** Unpublished data.
42. **Mackett, M., and G. L. Smith.** 1986. Vaccinia virus expression vectors. J. Gen. Virol. 67:2067–2082.
43. **Martell, M., J. I. Esteban, J. Quer, J. Genesca, A. Weiner, R. Esteban, J. Guardia, and J. Gomez.** 1992. Hepatitis C virus (HCV) circulates as a population of different but closely related genomes: quasispecies nature of the HCV genome distribution. J. Virol. 66:3225–3229.
44. **Mason, P. W.** 1989. Maturation of Japanese encephalitis virus glycoproteins produced by infected mammalian and mosquito cells. Virology 169:354–364.
45. **Matsuura, Y., S. Harada, R. Suzuki, Y. Watanabe, Y. Inoue, I. Saito, and T. Miyamura.** 1992. Expression of processed envelope protein of hepatitis C virus in mammalian and insect cells. J. Virol. 66:1425–1431.
46. **Meyers, G., N. Tautz, R. Stark, J. Brownlie, E. Dubovi, M. S. Collett, and H.-J. Thiel.** 1992. Rearrangement of viral sequences leads to a cytopathogenic pestivirus. Virology 191:368–386.
47. **Miller, R. H., and R. H. Purcell.** 1990. Hepatitis C virus shares amino acid sequence similarity with pestiviruses and flaviviruses as well as members of two plant virus supergroups. Proc. Natl. Acad. Sci. USA 87:2057–2061.
48. **Moss, B., O. Elroy-Stein, T. Mizukami, W. A. Alexander, and T. R. Fuerst.** 1990. New mammalian expression vectors. Nature (London) 348:91.
49. **Negro, F., D. Pacchioni, Y. Shimizu, R. H. Miller, G. Bussolati, R. H. Purcell, and F. Bonino.** 1992. Detection of intrahepatic replication of hepatitis C virus RNA by *in situ* hybridization and comparison with histopathology. Proc. Natl. Acad. Sci. USA 89:2247–2251.
50. **Ogata, N., H. J. Alter, R. H. Miller, and R. H. Purcell.** 1991. Nucleotide sequence and mutation rate of the H strain of hepatitis C virus. Proc. Natl. Acad. Sci. USA 88:3392–3396.
51. **Okamoto, H., M. Kojima, S.-I. Okada, H. Yoshizawa, H. Iizuka, T. Tanaka, E. E. Muchmore, D. A. Peterson, Y. Ito, and S. Mishiro.** 1992. Genetic drift of hepatitis C virus during an 8.2 year infection in a chimpanzee: variability and stability. Virology 190:894–899.
52. **Okamoto, H., S. Okada, Y. Sugiyama, K. Kurai, H. Iizuka, A. Machida, Y. Miyakawa, and M. Mayumi.** 1991. Nucleotide sequence of the genomic RNA of hepatitis C virus isolated from a human carrier: comparison with reported isolates for conserved and divergent regions. J. Gen. Virol. 72:2697–2704.
53. **Poch, O., I. Sauvaget, M. Delarue, and N. Tordo.** 1989. Identification of four conserved motifs among the RNA-dependent polymerase encoding elements. EMBO J. 8:3867–3874.
54. **Purcell, R. H., J. H. Walsh, P. V. Holland, A. G. Morrow, S. Wood, and R. M. Chanock.** 1971. Seroepidemiological studies of transfusion-associated hepatitis. J. Infect. Dis. 123:406–413.
55. **Rice, C. M., E. M. Lenches, S. R. Eddy, S. J. Shin, R. L. Sheets, and J. H. Strauss.** 1985. Nucleotide sequence of yellow fever virus: implications for flavivirus gene expression and evolution. Science 229:726–733.
56. **Rice, C. M., E. G. Strauss, and J. H. Strauss.** 1986. Structure of the flavivirus genome, p. 279–326. *In* S. Schlesinger and M. J. Schlesinger (ed.), The Togaviridae and Flaviviridae. Plenum Press, New York.
57. **Rice, C. M., and J. H. Strauss.** 1982. Association of Sindbis virion glycoproteins and their precursors. J. Mol. Biol. 154:325–348.
58. **Rice, C. M., and J. H. Strauss.** 1990. Production of flavivirus polypeptides by proteolytic processing. Sem. Virol. 1:357–367.
59. **Rümenapf, T., R. Stark, G. Meyers, and H.-J. Thiel.** 1991. Structural proteins of hog cholera virus expressed by vaccinia virus: further characterization and induction of protective immunity. J. Virol. 65:589–597.
60. **Saiki, R. K., D. H. Gelfand, S. Stoffel, S. J. Scharf, R. Higuchi, G. T. Horn, K. B. Mullis, and H. A. Erlich.** 1988. Primer-directed enzymatic amplification of DNA with a thermostable DNA polymerase. Science 239:487–491.
61. **Saito, I., T. Miyamura, A. Ohbayashi, H. Harada, T. Katayama, S. Kikuchi, Y. Watanabe, S. Koi, M. Onji, Y. Ohta, Q.-L. Choo, M. Houghton, and G. Kuo.** 1990. Hepatitis C virus infection is associated with the development of hepatocellular carcinoma. Proc. Natl. Acad. Sci. USA 87:6547–6549.
62. **Sambrook, J., E. F. Fritsch, and T. Maniatis.** 1989. Molecular cloning: a laboratory manual, 2nd ed. Cold Spring Harbor Laboratory, Cold Spring Harbor, N.Y.
63. **Schägger, H., and G. vonJagow.** 1987. Tricine-sodium dodecyl sulfate polyacrylamide gel electrophoresis for the separation of proteins in the range of 1 to 100 kDa. Anal. Biochem. 166:368–379.
64. **Spaete, R. R., D. Alexander, M. E. Rugroden, Q.-L. Choo, K. Berger, K. Crawford, C. Kuo, S. Leng, C. Lee, R. Ralston, K. Thudium, J. W. Tung, G. Kuo, and M. Houghton.** 1992. Char-

acterization of the hepatitis E2/NS1 gene product expressed in mammalian cells. Virology 188:819–830.

65. Strauss, E. G., and J. H. Strauss. 1985. Assembly of enveloped animal viruses, p. 205–234. *In* S. J. Casjens (ed.), Virus structure and assembly. Jones and Bartlett, Portola Valley, Calif.
66. Studier, F. W., A. H. Rosenberg, J. J. Dunn, and J. W. Dubendorff. 1990. Use of T7 RNA polymerase to direct expression of cloned genes. Methods Enzymol. 185:60–89.
67. Takamizawa, A., C. Mori, I. Fuke, S. Manabe, S. Murakami, J. Fujita, E. Onishi, T. Andoh, I. Yoshida, and H. Okayama. 1991. Structure and organization of the hepatitis C virus genome isolated from human carriers. J. Virol. 65:1105–1113.
68. Tanaka, T., N. Kato, M. Nakagawa, Y. Ootsuyama, M.-J. Cho, T. Nakazawa, M. Hijikata, Y. Ishimura, and K. Shimotohno. 1992. Molecular cloning of hepatitis C virus genome from a single Japanese carrier: sequence variation within the same individual and among infected individuals. Virus Res. 23:39–53.
69. Thiel, H.-J., R. Stark, E. Weiland, T. Rumenapf, and G. Meyers. 1991. Hog cholera virus: molecular composition of virions from a pestivirus. J. Virol. 65:4705–4712.
70. Weiland, E., R. Stark, B. Haas, T. Rümenapf, G. Meyers, and H.-J. Thiel. 1990. Pestivirus glycoprotein which induces neutralizing antibodies forms part of a disulfide-linked heterodimer. J. Virol. 64:3563–3569.
71. Weiner, A. J., M. J. Brauer, J. Rosenblatt, K. H. Richman, J. Tung, K. Crawford, F. Bonino, G. Saracco, Q.-L. Choo, M. Houghton, and J. H. Han. 1991. Variable and hypervariable domains are found in the regions of HCV corresponding to the flavivirus envelope and NS1 proteins and the pestivirus envelope glycoproteins. Virology 180:842–848.
72. Weiner, A. J., G. Kuo, D. W. Bradley, F. Bonino, G. Saracco, C. Lee, J. Rosenblatt, Q.-L. Choo, and M. Houghton. 1991. Detection of hepatitis C viral sequences in non-A, non-B hepatitis. Lancet 335:1–3.
73. Wengler, G., and G. Wengler. 1989. Cell-associated West Nile flavivirus is covered with E+pre-M protein heterodimers which are destroyed and reorganized by proteolytic cleavage during virus release. J. Virol. 63:2521–2526.
74. Wengler, G., and G. Wengler. 1991. The carboxy-terminal part of the NS3 protein of the West Nile flavivirus can be isolated as a soluble protein after proteolytic cleavage and represents an RNA-stimulated NTPase. Virology 184:707–715.
75. Wiskerchen, M., and M. S. Collett. 1991. Pestivirus gene expression: protein p80 of bovine viral diarrhea virus is a proteinase involved in polyprotein processing. Virology 184:341–350.

*Proc. Natl. Acad. Sci. USA*
Vol. 87, pp. 2057–2061, March 1990
Evolution

# Hepatitis C virus shares amino acid sequence similarity with pestiviruses and flaviviruses as well as members of two plant virus supergroups

(non-A, non-B hepatitis/potyvirus/carmovirus/picornavirus/alphavirus)

ROGER H. MILLER AND ROBERT H. PURCELL

Hepatitis Viruses Section, Laboratory of Infectious Diseases, National Institute of Allergy and Infectious Diseases, National Institutes of Health, Bethesda, MD 20892

*Contributed by Robert H. Purcell, December 27, 1989*

**ABSTRACT** Hepatitis C virus (HCV) is an important human pathogen that is associated with transfusion-related non-A, non-B hepatitis. Recently, HCV cDNA was cloned and the nucleotide sequence of approximately three-quarters of the virus genome was determined. A region of the predicted polyprotein sequence was found to share similarity with a nonstructural protein encoded by dengue virus, a member of the flavivirus family. We report here that HCV shares an even greater degree of protein sequence similarity with members of the pestivirus group (i.e., bovine viral diarrhea virus and hog cholera virus), which are thought to be distantly related to the flaviviruses. In addition, we find that HCV shares significant protein sequence similarity with the polyproteins encoded by members of the picornavirus-like and alphavirus-like plant virus supergroups. These data suggest that HCV may be evolutionarily related to both plant and animal viruses.

In recent years non-A, non-B (NANB) hepatitis has become the most common form of posttransfusion hepatitis (for reviews, see refs. 1–4). Although first discovered over a decade ago the etiological agent has remained elusive (5, 6). Studies involving the experimental inoculation of chimpanzees provided evidence that the infectious agent was a lipid-containing virus 30–60 nm in diameter bearing strong resemblance to members of the Togaviridae family (7–11). Since titer of the virus in serum rarely reaches $10^6$ chimpanzee infectious doses in patients, or experimentally infected animals, additional research has been difficult.

Recently, a λgt11 library was constructed with cDNA synthesized from the RNA of the putative etiological agent of NANB hepatitis (12). Protein synthesized by a specific recombinant reacted exclusively with sera from NANB patients (13). Molecular hybridization analysis demonstrated that the etiological agent, termed hepatitis C virus (HCV), is an RNA virus with a genome size of ≈10 kilobases. The sequence of nearly three-quarters of the virus genome has been reported (14). Analysis indicates that the virus genome is of the plus, or message sense, polarity and appears to lack a poly(A) tail at its 3′ end. The virus genome encodes a single polyprotein, a portion of which shares amino acid sequence similarity with the nonstructural number 3 (NS3) protein of dengue type 2 virus, a member of the flavivirus family. Additional computer-assisted protein analysis, presented here, demonstrates that HCV shares sequence similarity with the polyproteins of animal pestiviruses as well as those of the carmovirus and potyvirus families of plant viruses.

The publication costs of this article were defrayed in part by page charge payment. This article must therefore be hereby marked "*advertisement*" in accordance with 18 U.S.C. §1734 solely to indicate this fact.

## MATERIALS AND METHODS

**Computer Analysis.** Computer analysis was through the BIONET National Computer Resource for Molecular Biology. The program FASTA (15) was used to search the European Molecular Biology Organization (EMBO) and GenBank nucleotide data bases and the Swiss (SWS) and National Biomedical Research Foundation (NBRF) protein data bases for sequences with similarity to HCV sequences. FASTA, a derivative of the FASTP program that can be used for both nucleotide and amino acid data base searches, allows multiple regions of similarity between two sequences to be joined to determine a maximum alignment. Briefly, for a protein data base search, an initial similarity score is calculated based on a parameter that determines how many consecutive identities are required in a match and on the total number of identical and similar amino acids as specified by the PAM-250 matrix (16). Next, the FASTA program determines whether several regions with high initial similarity values can be aligned. If so, the program produces an optimal similarity score. There are several limitations imposed when using this program on BIONET. One is that only data base files, and not individual user files, can be analyzed. The second limitation is that only one scoring matrix (i.e., the PAM-250 matrix) can be used for the analysis. Within the FASTA program is a program RDF2 that evaluates the statistical significance of similarity scores by calculating a mean value and the standard deviation from the mean for the similarity scores of sequences in the data base. In this study, a stringent cutoff value for significance of ≥20% amino acid identity in ≥100 residues was also incorporated. Values cited in the text are given as optimized similarity scores with accompanying standard deviation units above the mean calculated for each data base search.

Three programs were used to determine regions of amino acid similarity considering only identical matches in the scoring matrix (17–19). The program HOMOLOGY was used to search for local regions of identity. Residues occurring in the alignments are cited in the text along with the probability that the matches occurred due to chance (e.g., $P = 0.05$ signifies that there is a 5% chance that the same match could occur between random sequences of the same size). The program ALIGN was used to determine the similarity over longer protein domains that encompassed regions with statistically significant matches of identical amino acids. The calculated value $H_{max}$ is directly proportional to the degree of similarity between two sequences over a region of defined size. It should be noted that $H_{max}$ scores produced by the alignment of random sequences range from 20 to 25 for sequences of 190

Abbreviations: HCV, hepatitis C virus; NANB, non-A, non-B; CARMV, carnation mottle virus; NS, nonstructural.

amino acids using the default parameter settings of the program and a segment size of 195 amino acids. Finally, the program GENALIGN was used for multiple sequence alignment.

## RESULTS

Houghton *et al.* (14) have reported the nucleotide sequence of approximately three-quarters of the HCV genome. The predicted polyprotein sequence, translated from the NS protein region of the HCV genome, is 2416 amino acids long. Analysis by Houghton and coworkers revealed that, among the virus sequences examined, the polyprotein sequence of HCV was most similar to that of a flavivirus. They reported a similarity between a 530-amino acid domain of the HCV polyprotein sequence and the NS3 protein sequence of dengue virus. We were intrigued by the uniqueness of the HCV sequence and performed searches using several programs to identify global or local regions of significant similarity between HCV and other sequences. This was of special interest since the nucleotide sequences of two pestivirus genomes, bovine viral diarrhea virus (20) and hog cholera virus (21), were determined recently.

First, we used computer-assisted nucleotide sequence analysis to look for similarity between HCV and any sequence recorded in the data base files. Computer searches conducted using the program FASTA with the HCV RNA genome as the query sequence did not result in a statistically significant match with nucleotide sequences in the EMBO or GenBank data bases. These results are in agreement with those of Houghton and coworkers (14). Thus, we conclude that the genome of HCV is not closely related to that of any known RNA virus.

Next, data base searches using the FASTA program and the PAM-250 matrix of Dayhoff (16) were performed to detect protein sequences possessing significant global similarity to the HCV polyprotein. HCV query sequences used were the complete 2416-amino acid polyprotein sequence, as well as the N terminus (i.e., residues 1–1299), and the C terminus (i.e., residues 1200–2416) of the reported HCV genome polyprotein. Searches were conducted using both the SWS and NBRF protein sequence data bases. The FASTA search of the NBRF data base using the entire 2416-residue HCV sequence produced one statistically significant alignment. We found that the amino acid sequence of HCV shared 20.6% amino acid identity with the dengue type 2 virus (22) NS3 protein over a 618-amino acid domain that encompassed the 530-amino acid region of similarity reported by Houghton *et al.* (14). In addition to the 141 matches between identical amino acids, there were 262 amino acids matched by the PAM-250 matrix for a total similarity of 60%. The optimized similarity score of 137 was 11.6 SD units away from the mean value of the analysis. The search of the SWS data base using the 2416-residue HCV polyprotein did not produce a statistically significant alignment. Therefore, using the 2416-amino acid sequence as the query sequence only one alignment score was statistically significant in our analysis.

The FASTA search of both the NBRF and SWS data bases with the N terminus of the HCV polyprotein as the query sequence yielded an alignment that was identical to the one described above. The FASTA search of the two data bases using the C terminus of the HCV polyprotein as the query sequence produced unexpected results. A statistically significant alignment was identified between residues 2058 and 2380 of the HCV polyprotein and the putative replicase of carnation mottle virus (CARMV), a member of the carmovirus group of plant viruses (23). Over a domain of 331 amino acids 67 (20%) of the residues were identical and 126 (38%) were scored as similar by the PAM-250 matrix for a total similarity of 58% (Fig. 1). The optimized similarity score of the alignment was 140, which was 11 SD units above the mean score of the search. Overall, the HCV polyprotein was found to possess significant global similarity to only two sequences in the protein data bases.

```
         2058
HCV      VRCHARKAVTHINSVWKDLLEDNVTPIDTTIMA--KNEVFCVQPEKGGRKPARLIVFPDL
         | |:: :  |  :::  :||: :::: |:::::  |:| | |: :::   ::     |
CARMV    VDCYQGRKRTIYENAAASLLDRAIERKDGDLKTFIKAEKFNVNLKSDPAPRVIQPRSPRY
         356

HCV      GVRVCEKMALYDVVTKLPLAVM--GSSYGFQYSPGQRVEFLVQAWKSKKTPMGFSYDTRC
         :| :   :  |:  :  :|: :  |::    |:::: :: :  ||:  :||:::::| :
CARMV    NVELGRYLKKYEHHAYKALDKIWGGPTVMKGYTTEEVAQHIWSAWNQFQTPVAIGFDMSR

HCV      FDSTVTESDIRTEEAIYQCCDLDPQARVAIKSLTERLYVGGPLTNSRGENCGYRRCRASG
         ||  |: :::  |:: |  | :: :|::|     : :  |  ::::   :::   || ||
CARMV    FDQHVSVAALEFEHSCYLAC-FEGDAHLANLLKMQLVNHGVGFASNGMLRYTKEGCRMSG
                                                                    *

HCV      VLTTSCGN-TLTCYIKARAACRAAGLQDCTMLVCGDDLVVICESAGVQEDAASLRAFTEA
          ::|: ||  |:| |: :          : ::  ||| |:|||:::::  :::|   |::
CARMV    DMNTALGNCLLACLITKHLM-----KIRSRLINNGDDCVLICERTDIDYVVSNL---TTG
            *   *                          ***

HCV      MTRYSAPPGDPPQPEYDLELITSCSSNVSVAHDGAGKRVYYLTRDPTTPLARAAWETAR-
          :|::     : :| |::| |  |  ::: : ||||   : ::||| ::::::: : ::
CARMV    WSRFGF-NCIAEEPVYEMEKIRFC--QMAPVFDGAG---WLMVRDPLVSMSKDSHSLVHW

                                     2380
HCV      --HTPVNSWLGNIIMFAPTLWARMILMTHFF
           :| :: || :: | :  : : : :: :|:
CARMV    NNETNAKQWLKSVGMCGLRIAGGVPVVQEFY
                                      671
```

FIG. 1. Alignment of the HCV polyprotein sequence (single-letter code) with the putative replicase of CARMV. Residues 2058–2380 of the predicted genome polyprotein of HCV (14) are aligned with residues 356–671 of CARMV (23) that are thought to represent the sequences specifying the virus replicase. Identical amino acid matches are connected with a solid line, while matches scored as similar by the PAM-250 matrix are connected with a colon. Dashes represent spaces between adjacent amino acids that have been inserted to optimize the alignment. Asterisks highlight the six amino acids that have been shown to be invariant among RNA virus replicases (24).

Next, we used several programs to determine whether the HCV polyprotein shared local regions of similarity with other virus sequences scoring only identical amino acid matches. Analysis using the program HOMOLOGY revealed the presence of statistically significant amino acid matches between HCV and two pestivirus polyprotein sequences. For example, the HCV sequences VVLATATPPGSVT (residues 874–886) and QRRGRTGRGKPGIYR (residues 1016–1030) were statistically similar to the bovine viral diarrhea virus (20) sequences VVAMTATPAGSVT (residues 2043–2055) and QRRGRVGRVKPGRYYR (residues 2199–2214) at the $P$ = 0.007 and 0.0005 levels, respectively. For reference purposes, we term the former HCV sequence region A and the latter HCV sequence region B. Similar findings were obtained when analyzing the hog cholera virus protein sequence (21). HCV regions A and B were also found to be similar to flavivirus and plant potyvirus polyprotein sequences; however, no such similarity was detected by comparing HCV to alphavirus, rubivirus, or picornavirus protein sequences. For example, the HCV sequence TATPPGS (residues 878–884) in region A was found to be identical to the dengue type 4 virus (25) sequence TATPPGS (residues 1796–1802), which is a statistically significant match at the $P$ = 0.044 level. This sequence alignment was also present in the global alignment of Houghton and coworkers (14) and in our alignment using the program FASTA as described above. In addition, the HCV sequence LVVLATATPPG (residues 873–883) of region A was significantly similar to the tickborne encephalitis virus NS3 sequence (26) LVLMTATPPG (residues 1806–1815) at the $P$ = 0.019 level of significance. Significant similarity was also found between HCV sequence region B and a plant potyvirus protein sequence. Specifically, the HCV sequence QRRGRTGRGKPG (residues 1016–1027) was similar to the sequence QRFGRVGRNKPG (residues 1463–1474) of the tobacco vein mottling virus (27) at the $P$ = 0.018 level of significance. Overall, two regions of the "NS3-like" region of the HCV polyprotein were found to share sequence similarity with pestivirus, flavivirus, and potyvirus proteins.

To determine the degree of relatedness among HCV and the proteins of the pesti-, flavi-, and potyviruses, we used several programs to analyze a 190-residue domain encompassing HCV regions A and B. In the program ALIGN, the calculated value $H_{max}$ is directly proportional to the degree of similarity between two sequences over a region of defined size. The analysis indicated that the 190-amino acid region of HCV was most similar to that of bovine viral diarrhea virus ($H_{max}$ = 52), hog cholera virus ($H_{max}$ = 51), and tobacco vein mottling virus ($H_{max}$ = 47). Interestingly, HCV shared more similarity with the potyvirus sequence than it did with any of the flavivirus sequences ($H_{max}$ = 33–41) examined (Table 1). Multiple sequence alignment of these four sequences using the program GENALIGN demonstrates that there are 25 amino

Table 1. $H_{max}$ similarity values

| | HCV | HOG | BVD | TBE | JEV | YFV | DEN | WNF | KUN | TVM |
|---|---|---|---|---|---|---|---|---|---|---|
| HCV | — | 51 | 52 | 41 | 41 | 40 | 38 | 37 | 33 | 47 |
| HOG | — | — | 169 | 38 | 34 | 33 | 35 | 37 | 33 | 47 |
| BVD | — | — | — | 39 | 34 | 34 | 38 | 36 | 37 | 48 |
| TBE | — | — | — | — | 91 | 87 | 90 | 90 | 93 | 35 |
| JEV | — | — | — | — | — | 88 | 118 | 159 | 163 | 35 |
| YFV | — | — | — | — | — | — | 94 | 95 | 90 | 39 |
| DEN | — | — | — | — | — | — | — | 117 | 121 | 36 |
| WNF | — | — | — | — | — | — | — | — | 177 | 34 |
| KUN | — | — | — | — | — | — | — | — | — | 35 |
| TVM | — | — | — | — | — | — | — | — | — | — |

The following virus sequences were used in the analysis: HCV (14); HOG, hog cholera virus (24); BVD, bovine viral diarrhea virus (23); TBE, tickborne encephalitis virus (25); JEV, Japanese encephalitis virus (28); YFV, yellow fever virus (29); DEN, dengue virus (25); WNF, West Nile fever virus (30); KUN, Kunjin virus (31); TVM, tobacco vein mottling virus (26).

```
HCV     VVlaTATPPGSVT---vpHPnIEE--valsttGE------ipfyGkaIPleviKGgrhLiFchskkkc
        ||  |||| ||||     || |||        ||          |  ||    ||   | |
BVD     VVAMTATPAGSVTTTGQKHP-IEEFIAPEVMKGEDLGSqfLDIAGLKIPVdEMKG-NMLVFVPTRNMA
        |||||||||| ||||||||| |||||||||||||||||  |||||||||| |||  ||||||||||||
HOG     VVAMTATPAGtVTTTGQKHP-IEEFIAPEVMKGEDLGSeyLDIAGLKIPVeEMK-nNMLVFVPTRNMA
             ||| |            |  |               |
TVM     iikvsATPPGrecdltt-qypvEllIeeqlslrdfvdaqgtDahadvvkkgdnilvyvasynevdqls
             *** *            *

HCV     dELAaKLvAlGiNavaYY--rGldvsvipTS-gdvVvVATdAlmtGyT-gDfDsVID-cntCvtqtVd
         | | || | | |   ||           ||    | ||| |   | |  | | |||    |
BVD     VEvAKKLKAKGYNSGYYYSGEDPaNLRVVTSQSPYViVATNAIESGVTLPDLDtVIDTGLKCEKRvRv
        || |||||||||||||||||||| ||||| |||||| |||||||||||||||| | ||||||||| |
HOG     VEaAKKLKAKGYNSGYYYSGEDPsNLRVVaSQSPYVvVATNAIESGVTLPDLDVVVDTGLKCEKRiRl
                      |               |      |||| || |||| | ||||| |||
TVM     kmLnergflvtkvdGrtmklgGveiitkgsSikkhfiVATNiIEnGVTL-DvDVVVDfGLK----vVp
                                      *      ***     * *  * * * *

HCV     fSldPtftietitlPqdavsRtQRRGRtGRgKPGi-YR
         |  |                 ||||| || |||  ||
BVD     SSKiPFIVTGLKRMAVTvGEQAQRRGRVGRVKPGRYYR
        | | ||||||||||||| ||||||||||||||||||||
HOG     SpKmPFIVTGLKRMAVTiGEQAQRRGRVGRVKPGRYYR
                          ||  || ||||| |||   |
TVM     nldsdnrlvsyckiPislGERiQRfGRVGRnKPGvalR
                              ** ** ** ***   *
```

FIG. 2. Multiple sequence alignment of a conserved domain in the genome proteins (single-letter code) of HCV, pestiviruses, and a plant potyvirus. Alignment of the following regions of the genome polyproteins of four viruses are shown: HCV, residues 874–1030 of HCV (14); BVD, residues 2025–2196 of bovine diarrhea virus (20); HOG, residues 1886–2057 of hog cholera virus (21); TVM, residues 1311–1477 of tobacco vein mottling virus (27). Identically matched amino acids between two or more virus proteins are shown as capital letters connected with a straight line. Unmatched amino acids are depicted with lowercase letters. Dashes represent spaces between adjacent amino acids that have been inserted to optimize the alignment. Invariant residues are highlighted with an asterisk.

acids that are invariant among these diverse virus proteins (Fig. 2). Thus, it is likely that this region was conserved in evolution because the protein has an important biological function in virus replication or gene expression.

## DISCUSSION

In this study, we used computer-assisted protein analysis to search for sequences with significant similarity to the HCV polyprotein. To identify sequences sharing global similarity, we used a data base searching program that incorporated the PAM-250 matrix to produce alignments consisting of identical and similar amino acid matches. The analysis revealed that the HCV polyprotein possessed statistically significant similarity to only two sequences in the protein data bases. Both sequences were viral in origin. First, the NS3 protein of dengue type 2 virus aligned with a 618-residue domain located near the N terminus of the HCV polyprotein. This represents an extension of nearly 100 amino acids over an alignment reported by Houghton and coworkers (14) that spanned 530 residues within the same region. Second, the putative replicase of CARMV aligned with a region at the C terminus of the HCV polyprotein. This finding was unexpected since CARMV, a member of the carmovirus family, is a plant virus. Overall, the polyprotein of HCV was found to share global similarity with protein sequences encoded by RNA viruses of both animals and plants, which adds support to the hypothesis that there is an evolutionary relationship between these two virus groups.

Analysis in which programs were used to search for regions of local identity of amino acids revealed that regions of the HCV polyprotein aligned with the NS3 protein sequence of flaviviruses and with corresponding regions of the polyproteins of pestiviruses and plant potyviruses. The similarity was the greatest between HCV and pestiviruses. The reason that this similarity was not detected by others previously, or in our data base searches, was that the pestivirus sequences were published only recently and were not in the data bases for analysis. (Therefore, we analyzed the sequences from user files that we created.) Unexpectedly, we did not find significant similarity between the HCV genome protein sequence and the putative replicase of the flaviviruses or pestiviruses.

Comparative analysis of the polyproteins of the members of the flavivirus family reveals that the sequences of the NS proteins are highly conserved (Fig. 3). Multiple sequence alignment of the predicted polyprotein sequences of Japanese encephalitis (28), yellow fever (29), West Nile (30), Kunjin (31), tickborne encephalitis (26), and three dengue virus isolates (25, 32, 33) demonstrates that there are several regions of high amino acid conservation. Within the consensus polyprotein sequence of ≈3400 amino acids there are 21 domains that possess 5 or more consecutive amino acids that are identical in every flavivirus sequence (unpublished data). Eight of these domains are located in the NS3 protein sequence. The 190-amino acid domain of NS3 that shares sequence similarity with HCV contains 3 of these conserved domains. The first is a 7-residue sequence MTATPPG found at the N terminus of the domain. The second is a 5-residue sequence EMGAN near the C terminus. The third is an 8-residue sequence SAAQRRGR located at the extreme C terminus of the domain. Regarding the latter sequence, although the next 3′ residue is variable among flavivirus sequences the following 2 residues are always GR. Our

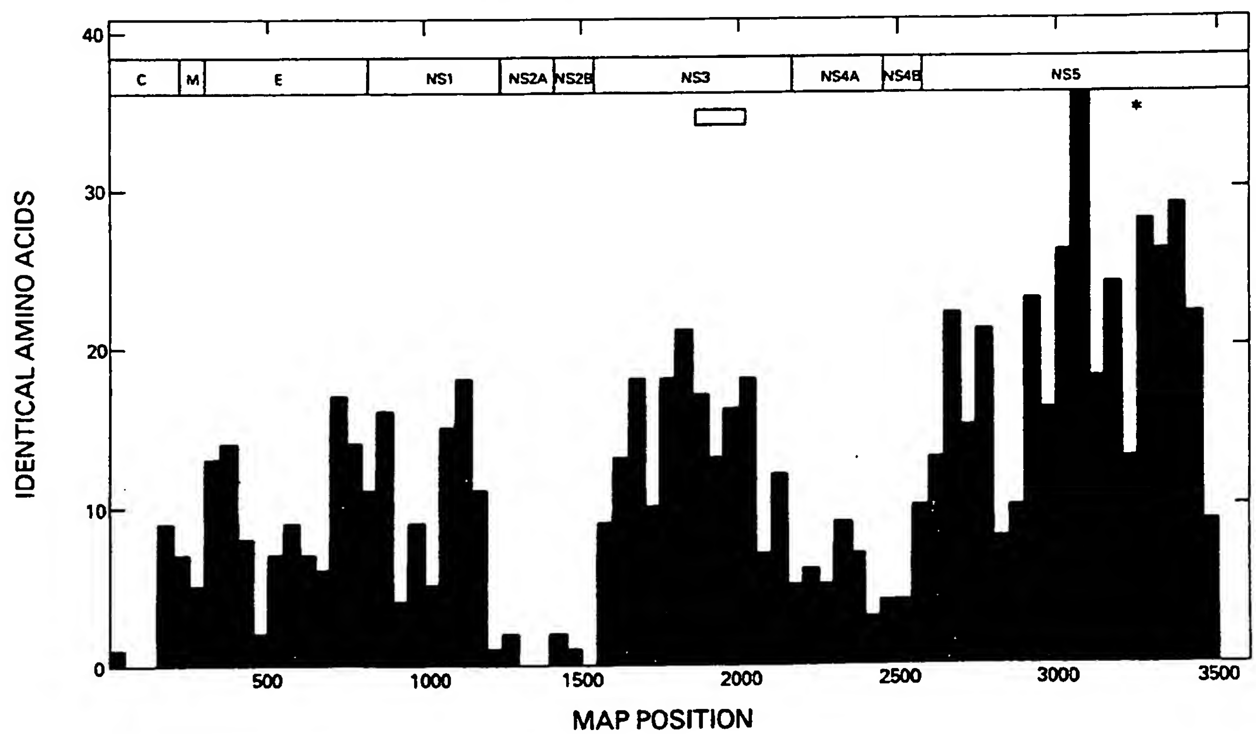



FIG. 3. Histogram of invariant amino acids in the genome polyprotein of the flaviviruses. The program GENALIGN was used to align the amino acids of the following flaviviruses: three isolates of dengue virus (25, 32, 33), Kunjin virus (31), Japanese encephalitis virus (28), tickborne encephalitis virus (26), West Nile virus (30), and yellow fever virus (29). The number of identical amino acids at each position for all 8 sequences, within a block of 50 contiguous residues, is plotted against the position of the residues on the consensus genome polyprotein. The insertion of gaps to optimize the alignment resulted in a total length of the consensus sequence that was longer than any of the individual polyproteins. The gene order of the polyprotein is shown at the top illustrating the position of the structural proteins [i.e., the capsid (C), matrix (M), and envelope (E) proteins] and the NS proteins. The open box under the NS3 protein heading depicts the 190-amino acid domain that shares sequence similarity with regions A and B of the HCV polyprotein. The asterisk represents the position of the invariant GDD moiety of RNA virus replicases.

analysis indicates that only the first and third domains share significant similarity to HCV in the regions of the polyprotein sequence that we have termed A and B.

The NS3 gene region of flaviviruses may encode a protein with several enzymatic activities. First, the N terminus of the NS3 protein is known to share sequence similarity with serine proteases (34). Second, the central domain of NS3 of both flaviviruses and plant potyviruses has been shown to share sequence similarity with helicase-like nucleoside triphosphate binding (NTB) proteins from eukaryotic and prokaryotic cells (35). We find that HCV also shares similarity to NTB proteins in regions A and B of the polyprotein sequence (unpublished data). Thus, it is possible that flavi-, poty-, and pestiviruses, as well as HCV, encode a NTB protein that has been conserved in evolution because of its important catalytic function in virus gene expression or replication.

The NS5 protein has the most highly conserved amino acid sequence of any of the flavivirus proteins and is thought to encode the virus replicase. Within NS5 there are 10 domains that contain $\geq 5$ consecutive identical amino acids including the longest tract of invariant residues (i.e., 14 amino acids) identified in the alignment of the polyproteins. In addition, all flavivirus NS5 proteins possess the 6-amino acid residues that are known to be invariant among RNA polymerase sequences (24). Despite the fact that NS5 is more highly conserved than NS3, we found that there was no statistically significant similarity between the flavivirus NS5 protein and the HCV polyprotein using global or local alignment programs. The only sequence that possessed statistically significant similarity with a region at the C terminus of the HCV polyprotein sequence was the putative replicase of CARMV. Therefore, the HCV replicase may be most closely related to that of a plant virus.

Overall, we find that HCV sequences share significant similarity with proteins from members of two unrelated plant virus families. RNA viruses of plants have been assigned to two supergroups based on the similarity of their genome and protein sequences to either the picorna- or the alphaviruses of animals. The picornavirus supergroup consists of the como-, nepo-, and potyviruses, while the alphavirus, or Sindbis-like, supergroup consists of the alfalfa mosaic, ilar-, bromo-, cucumo-, tobamo-, potex-, tobra-, furo-, nordei-, tombus-, and carmovirus groups (36). There is some speculation that the tombusviruses and carmoviruses may belong to a third supergroup because of their unusually small genome size. The genome of the latter virus group is $\approx 4000$ nucleotides and does not encode an NS3-like protein. Our analysis indicates that amino acid sequences near the N terminus of the HCV polyprotein are similar to those of the potyviruses, while amino acid sequences near the C terminus of the HCV polyprotein are most similar to those of the carmoviruses. Thus, it is possible that HCV represents a recombinant virus possessing an N terminus derived from a picornavirus-like ancestor and a C terminus derived from an alphavirus-like ancestor. However, it is clear that HCV is not closely related to any of these RNA virus families or any other RNA virus family thus far described.

In conclusion, taxonomic classification of HCV must await analysis of the complete nucleotide sequence, which includes the genes encoding the structural proteins as well as the 5′ and 3′ noncoding regions. The data presented here suggest that HCV is distantly related to the pestiviruses and flaviviruses of animals and to members of two plant virus supergroups. It is possible that HCV is a recombinant virus since RNA recombination has been demonstrated for positive-strand (37) and negative-strand RNA viruses (38). Another possibility is that a single virus gave rise to HCV and these similar viruses. Thus, HCV may represent an evolutionary link between the plant virus supergroups and between viruses infecting both plants and animals.

We thank M. Collett and P. Mason for providing pestivirus and flavivirus sequences, respectively, that were not in the data base at the time of the analysis and M. Brinton for helpful discussions. We also thank N. Lofgren, D. Wong, T. Chestnut, and M. Lewis for help in entering nucleotide sequences into computer files and R. Chanock, M. Collett, and M. Brinton for comments on the manuscript. Computer-assisted nucleic acid and protein analysis was through the BIONET National Computer Resource for Molecular Biology, which is supported by National Institutes of Health Grant U41-01685-05.

1. Blum, H. E. & Vyas, G. N. (1982) *Haematologia* 15, 153–173.
2. Dienstag, J. L. (1983) *Gastroenterology* 85, 439–462.
3. Fagan, E. A. & Williams, R. (1984) *Semin. Liver Dis.* 4, 314–335.
4. Mattsson, L. (1989) *Scand. J. Infect. Dis. Suppl.* 59, 1–55.
5. Feinstone, S. M., Kapikian, A. Z. & Purcell, R. H. (1975) *N. Engl. J. Med.* 292, 767–770.
6. Alter, H. J., Purcell, R. H., Holland, P. V. & Popper, H. (1978) *Lancet* i, 459–463.
7. Feinstone, S. M., Alter, H. J., Dienes, H. P., Shimizu, Y., Popper, H., Blackmore, D., Sly, D., London, W. T. & Purcell, R. H. (1981) *J. Infect. Dis.* 144, 588–598.
8. Bradley, D. W., McCaustland, K. A., Cook, E. H., Schable, C. A., Ebert, J. W. & Maynard, J. E. (1985) *Gastroenterology* 88, 773–779.
9. Purcell, R. H., Gerin, J. L., Popper, H., London, W. T., Cicmanec, J., Eichberg, J. W., Newman, J. & Hrinda, M. E. (1985) *Hepatology* 5, 1091–1099.
10. He, L.-F., Alling, D., Popkin, T., Shapiro, M., Alter, H. J. & Purcell, R. H. (1987) *J. Infect. Dis.* 156, 636–640.
11. Fagan, E. A., Ellis, D. S., Tovey, G. M., Lloyd, G., Portman, B., Williams, R. & Zuckerman, A. J. (1989) *J. Med. Virol.* 28, 150–155.
12. Choo, Q.-L., Kuo, G., Weiner, A. J., Overby, L. R., Bradley, D. W. & Houghton, M. (1989) *Science* 244, 359–362.
13. Kuo, G., Choo, Q.-L., Alter, H. J., Gitnick, G. L., Redeker, A. G., Purcell, R. H., Miyamura, T., Dienstag, J. L., Alter, M. J., Stevens, C. E., Tegtmeier, G. E., Bonino, F., Colombo, M., Lee, W.-S., Kuo, C., Berger, K., Shuster, J. R., Overby, L. R., Bradley, D. W. & Houghton, M. (1989) *Science* 244, 362–364.
14. Houghton, M., Choo, Q.-L. & Kuo, G. (1988) *Eur. Patent Appl.* 88,310,922.5 and Publ. 318,216.
15. Pearson, W. R. & Lipman, D. J. (1988) *Proc. Natl. Acad. Sci. USA* 85, 2444–2448.
16. Dayhoff, M., Schwartz, R. M. & Orcutt, B. C. (1978) in *Atlas of Protein Sequence and Structure*, ed. Dayhoff, M. (Natl. Biomed. Res. Found., Silver Spring, MD), Vol. 5, Suppl. 3, pp. 345–352.
17. Needleman, S. B. & Wunsch, C. D. (1970) *J. Mol. Biol.* 48, 443–453.
18. Korn, L. J., Queen, C. L. & Wegman, M. N. (1977) *Proc. Natl. Acad. Sci. USA* 74, 4401–4405.
19. Brutlag, D. L., Clayton, J., Friedland, P. & Kedes, L. H. (1982) *Nucleic Acids Res.* 10, 279–294.
20. Collett, M. S., Larson, R., Gold, C., Strick, D., Anderson, D. K. & Purchio, A. F. (1988) *Virology* 165, 191–199.
21. Meyers, G., Rumenapf, T. & Thiel, H. (1989) *Virology* 171, 555–567.
22. Yaegashi, T., Vakharia, V. N., Page, K., Sasaguri, Y., Feighny, R. & Padmanabhan, R. (1986) *Gene* 46, 257–267.
23. Guilley, H., Carrington, J. C., Balazs, E., Jonard, G., Richards, K. & Morris, T. J. (1985) *Nucleic Acids Res.* 13, 6663–6677.
24. Domier, L. L., Shaw, J. G. & Rhoads, R. E. (1987) *Virology* 158, 20–27.
25. Mackow, E., Makino, Y., Zhao, B., Zhang, Y.-M., Markoff, L., Buckler-White, A., Guiler, M., Chanock, R. M. & Lai, C.-J. (1987) *Virology* 159, 217–228.
26. Mandl, C. W., Heinz, F. X., Stockl, E. & Kunz, C. (1989) *Virology* 173, 291–301.
27. Domier, L. L., Franklin, K. M., Shahabuddin, M., Hellman, G. M., Overmeyer, J. H., Hiremath, S. T., Siaw, M. F. E., Lomonossoff, G. P., Shaw, J. G. & Rhoads, R. E. (1986) *Nucleic Acids Res.* 14, 5417–5430.
28. Sumiyoshi, H., Mori, C., Fuke, I., Morita, K., Kuhara, S., Kondou, J., Kikuchi, Y., Nagamatu, H. & Igarashi, A. (1987) *Virology* 161, 497–510.
29. Rice, C. M., Lenches, E. M., Eddy, S. R., Shin, S. J., Sheets, R. L. & Strauss, J. H. (1985) *Science* 229, 726–733.
30. Castle, E., Leidner, U., Nowak, T. & Wengler, G. (1986) *Virology* 149, 10–26.
31. Coia, G., Parker, M. D., Speight, G., Byrne, M. E. & Westaway, E. G. (1988) *J. Gen. Virol.* 69, 1–21.
32. Deubel, V., Kinney, R. M. & Trent, D. W. (1988) *Virology* 165, 234–244.
33. Hahn, Y. S., Galler, R., Hunkapiller, T., Dalrymple, J. M., Strauss, J. H. & Strauss, E. G. (1988) *Virology* 162, 167–180.
34. Gorbalenya, A. E., Donchenko, A. P., Koonin, E. V. & Blinov, V. M. (1989) *Nucleic Acids Res.* 17, 3889–3897.
35. Lain, S., Riechmann, J. L., Martin, M. T. & Garcia, J. A. (1989) *Gene* 82, 357–362.
36. Goldbach, R. & Wellink, J. (1988) *Intervirology* 29, 260–267.
37. Cooper, P. D. (1965) *Virology* 25, 431–438.
38. Khatchikian, D., Orlich, M. & Rott, R. (1989) *Nature (London)* 340, 156–157.

JOURNAL OF VIROLOGY, July 1993, p. 4017–4026 Vol. 67, No. 7
0022-538X/93/074017-10$02.00/0


# NS3 Is a Serine Protease Required for Processing of Hepatitis C Virus Polyprotein

LICIA TOMEI, CRISTINA FAILLA, ELISA SANTOLINI, RAFFAELE DE FRANCESCO, AND NICOLA LA MONICA*

*Istituto di Ricerche di Biologia Molecolare P. Angeletti, Pomezia, 00040 Rome, Italy*

Received 22 January 1993/Accepted 5 April 1993

**Hepatitis C virus (HCV) possesses a positive-sense RNA genome which encodes a large polyprotein of 3,010 amino acids. Previous data and sequence analysis have indicated that this polyprotein is processed by cellular proteases and possibly by a virally encoded serine protease localized in the N-terminal domain of nonstructural protein NS3. To characterize the molecular aspects of HCV protein biogenesis and to clearly identify the protein products derived from the HCV genome, we have examined HCV polyprotein expression by using the vaccinia virus T7 transient expression system in transfected cells and by cell-free translation studies. HCV proteins were identified by immunoprecipitation with region-specific antisera. Here we show that the amino-terminal region of the HCV polyprotein is processed in vitro by cellular proteases releasing three structural proteins: p21 (core), gp37 (E1), and gp61 (E2). Processing of the nonstructural region of HCV was evident in transfected cells. Two proteins of 24 and 68 kDa were immunoprecipitated with anti-NS2 and NS3 antisera, respectively. Antiserum against NS4 recognized three proteins of 6, 26, and 31 kDa, while antisera specific for NS5 immunoprecipitated two polypeptides of 56 and 65 kDa, indicating that each of these two genes encodes at least two different proteins. When the NS3 protease domain was inactivated by replacing the proposed catalytic Ser-1165 with Ala, processing at several sites was abolished. When Ser-1164 was mutated to Ala, no effect on the processing was observed. Cleavage activities at three of the four sites affected by NS3 were shown to occur in *trans*, while processing at the carboxy terminus of NS3 could not be mediated in *trans*. These results provide a detailed description of the protein products obtained from the processing of the HCV polyprotein. Furthermore, the data obtained implicate NS3 as a serine protease and demonstrate that a catalytically active NS3 is necessary for cleavage of the nonstructural region of HCV.**

Hepatitis C virus (HCV) is considered to be the major etiologic agent of posttransfusion non-A, non-B hepatitis (4, 20). The enveloped virion consists of unknown species of structural proteins encoded by a positive-sense RNA genome. Sequence analysis has indicated that the viral genome is approximately 9,400 nucleotides long and includes a 5′ untranslated region of 341 nucleotides which precedes a single open reading frame (ORF) encoding a precursor polyprotein of 3,010 or 3,011 amino acids. This long ORF is followed by an untranslated region of 23 to 54 nucleotides located at the 3′ end (5, 18, 31). The genetic organization of the viral genome is similar to that of flaviviruses and pestiviruses, with the putative structural proteins located in the N-terminal region and a variety of nonstructural proteins located at the C terminus of the polyprotein (23). The putative structural region of HCV is shorter than that of flaviviruses and pestiviruses, and it lacks primary sequence similarity with these two virus families (6). However, it is organized in a similar fashion, with a basic N-terminal (p20) presumed nucleocapsid core protein (C), followed by two glycoproteins, gp35 and gp70. gp35 probably corresponds to the matrix/envelope glycoprotein in the virion (E1), whereas gp70 may correspond to an envelope glycoprotein equivalent to gp53/55 of pestiviruses (E2) or to the NS1 glycoprotein of flaviviruses (7, 25, 33). In vitro protein synthesis followed by amino acid sequence analysis of the products has demonstrated that these proteins are released from the precursor polyprotein by cellular proteases in association with membranes of the endoplasmic reticulum (14).

The nonstructural region of the HCV genome has not been characterized in detail, but it is thought to be processed in a manner similar to that of flaviviruses and pestiviruses, releasing a series of proteins from the polyprotein precursor. The exact number of processed protein products derived from this region has not been identified. However, by analogy with the flaviviruses, the putative nonstructural polypeptides of HCV have been called NS2, NS3, NS4, and NS5. Although the amino acid sequence of the HCV polyprotein differs from that of the flavivirus polyprotein, the two polyproteins have similar hydropathy profiles. Tentative boundaries of the nonstructural proteins of HCV have been assigned on the basis of this similarity (5, 18, 31, 32). The NS2 and NS4 proteins are very hydrophobic, probably membrane bound, and of unknown function. Their predicted molecular sizes are 25 and 52 kDa, respectively (16). In flaviviruses, two proteins are encoded by each of the NS2 and NS4 genes (NS2a+b, and NS4a+b) (2). The NS5 gene of HCV is predicted to encode a polypeptide of 116 kDa, and it contains a GDD consensus sequence found in several viral RNA-dependent RNA polymerases, suggesting that it may be involved in viral replication (7, 31).

Cleavages generating the N termini of the flavivirus nonstructural proteins NS2b, NS3, NS4a, and NS5 follow dibasic amino acid residues, occur rapidly and efficiently in infected cells, and are mediated by a viral protease located in the cytoplasm (2, 24). The NS3 protein of pestiviruses and flaviviruses was found to be a component of the viral protease, as determined from sequence analysis and molecular modeling studies. The positions of three amino acid residues (His-53, Asp-77, and Ser-138) located within the N-terminal domain of NS3 are strictly conserved among

* Corresponding author.

flaviviruses and are predicted to correspond spatially to the catalytic triad of trypsin-like serine proteases. Results from site-directed mutagenesis of these amino acid residues are consistent with the hypothesis that the catalytic activity resides in the NS3 domain and that these residues comprise the catalytic triad (1, 3, 34). Analysis of the amino acid sequence of the NS3 protein of HCV has suggested that this viral protein may encode a trypsin-like serine protease which could function in the processing of the viral polyprotein as in the case of flaviviruses and pestiviruses (23). Residues His-1083, Asp-1107, and Ser-1165, numbered according to their location within the HCV polyprotein, are found in the N-terminal domain of the NS3 protein and are highly conserved among all HCV strains sequenced so far (31). These residues are predicted to correspond spatially to the catalytic triad of the putative serine protease of HCV, suggesting that this protein may play a pivotal role in polyprotein processing. Furthermore, the NS3 polypeptide with a predicted molecular size of 69 kDa contains a nucleoside triphosphate-binding helicase domain that is presumably involved in unwinding of the RNA genome (3, 5).

In this study, we have examined HCV polyprotein expression and processing by cell-free protein synthesis studies and transient expression in transfected cells. We have identified specific protein products expressed by different regions of the viral genome by using region-specific antisera. We have obtained evidence that the structural components of HCV are processed in a fashion independent of NS3 synthesis, while a catalytically active NS3 protein is necessary for the correct maturation of the nonstructural proteins of HCV.

## MATERIALS AND METHODS

**Cells and virus.** HeLa cells, originally obtained from the American Type Culture Collection, were grown in Dulbecco's modified Eagle's essential minimal medium (MEM) containing 10% fetal calf serum (FCS). Vaccinia virus vTF7-3 (10) was grown in RK-13 cells plated in Eagle's MEM containing 10% FCS.

**Construction of recombinant plasmids.** Plasmid pCD(38-9.4) encodes the HCV sequences from nucleotides 1 to 9416 downstream of a T7 promoter. The clone was constructed by joining individual cDNA fragments derived from plasmids BK146, BK144, BK112.1, BK112.5, and BK166 at overlapping restriction sites (31). The HCV cDNA clone was introduced into the plasmid vector pCDNA-1 (Invitrogen). The cDNA subclones were provided by Hiroto Okayama (Osaka University) and represent HCV clones isolated from Japanese patients.

pCITE(146) is derived from clone BK146 and contains HCV sequences from an *Msc*I site engineered at nucleotide 333 to an *Xba*I site introduced at nucleotide 4840. The cDNA fragment derived by polymerase chain reaction (PCR) amplification with sequence-specific primers lacks its own ATG and the untranslated region of HCV. The fragment was cloned downstream of a T7 promoter in the pCITE vector (Novagen) and inserted downstream of the 5' untranslated region of encephalomyocarditis virus.

pSK(CORE) was derived from PCR amplification of nucleotides 330 to 938 (amino acid residues 1 to 200) with sequence-specific primers. The amplified DNA fragment derived from plasmid pCD(38-9.4) contains an *Xba*I site and an *Xho*I site engineered at the 5' end and the 3' end, respectively. The cDNA fragment was cloned downstream of a T7 promoter in the pBluescript vector SK II; it lacks the 5' untranslated region of HCV and encodes the nucleocapsid protein core (11).

To construct plasmid pCITE(SX), clone pCD(38-9.4) was cleaved with *Sac*II and *Xba*I, and a DNA fragment containing nucleotides 3303 to 9416 (amino acid residues 991 to 3010) was purified. The DNA fragment was then inserted downstream of the T7 promoter into the *Bst*XI and *Xba*I sites of the expression vector pCITE.

pCITE(NS4-5) was obtained by cloning into the *Bst*XI and *Stu*I sites of pCITE a *Sph*I-*Kpn*I fragment derived from pCD(38-9.4) after treatment of both DNAs with Klenow polymerase. The construct encodes nucleotides 5281 to 9071 (amino acid residues 1651 to 2921) downstream of the T7 promoter and of the 5' untranslated region of encephalomyocarditis virus.

pCITE(NS3) was derived from PCR amplification of nucleotides 3351 to 5171 (amino acid residues 1007 to 1616) with sequence-specific primers, using plasmid pCD(38-9.4) as the template. The amplified DNA fragment was cloned by blunt-end ligation into the expression vector pCITE which had been cleaved with *Nco*I and *Stu*I and blunted with Klenow polymerase.

Site-directed mutants in the NS3 catalytic serine, pCD(38-9.4:$S_{1165}$-A) and pCITE(SX$S_{1165}$-A), and the respective negative control mutants in the adjacent serine, pCD(38-9.4:$S_{1164}$-A) and pCITE(SX$S_{1164}$-A), were obtained by inserting the mutations in PCR primers that were then used to generate mutant DNA fragments according to the procedures of Higuchi et al. (13). The mutant DNA fragments were recloned into the parent plasmids by using restriction sites flanking the mutations and were subsequently sequenced. The triplet coding for serine 1165, TCG, was replaced by GCG, and that coding for serine 1164, TCT, was replaced by GCT. Both of these triplets code for alanine.

Constructs for the expression of TrpE fusion proteins with E1, E2/NS1, NS4, and NS5b sequences were made by using pATH plasmids (30). The NS2, NS3, and NS5a fusion proteins with glutathione *S*-transferase (GST) were made by using plasmid pGEX-3x (28). Cloning of the HCV fragments in the expression vectors was achieved by PCR amplification of the area of interest, using synthetic oligonucleotides containing appropriate restriction sites or by in-frame fusion of cDNA fragments by means of standard recombinant DNA protocols. Recombinant plasmids were transformed in *Escherichia coli* DH5α, with the exception of pCD(38-9.4), which was transformed in MC1061/P3.

**Induction of expression plasmids and preparation of fusion proteins.** TrpE fusion proteins were induced in *E. coli* DH5α cells harboring recombinant plasmids with of 3-β-hydroxyindoleacrylic acid at a final concentration of 5 ng/ml. GST fusion proteins were expressed in *E. coli* DH5α cells upon induction with 0.4 mM isopropyl-β-D-thiogalactopyranoside (IPTG). The TrpE-E1, TrpE-E2/NS1, GST-NS2, TrpE-NS4, and TrpE-NS5b fusion proteins accumulated in the inclusion bodies of *E. coli* and were prepared by lysis of bacteria, DNase I digestion, and precipitation of the insoluble fraction as described previously (30).

The GST-NS3 and GST-NS5a fusion proteins were in the soluble fraction. These proteins were affinity purified on a glutathione-Sepharose CL4B column (Pharmacia), the HCV portion of the GST-NS3 protein was cleaved from the fusion protein by factor Xa (New England Biolabs) and purified as described previously (28).

Sodium dodecyl sulfate (SDS)-polyacrylamide gel slices containing fusion proteins were ground in phosphate-buffered saline (PBS), emulsified in Freund's adjuvant, and used

to immunize rabbits (12). New Zealand White male and female rabbits were used for production of all antisera.

**Immunoaffinity purification of anti-HCV antibodies from patient sera.** Human antibodies against E1 and NS5a were immunopurified from patient sera. The TrpE-E1 protein was prepared in *E. coli* as described above, solubilized in SDS-polyacrylamide gel electrophoresis (PAGE) loading buffer, and run on a 10% polyacrylamide–SDS gel. The protein was transferred from the gel onto a nitrocellulose filter by electroblotting and used for immunoaffinity purification of anti-E1 antibodies as described previously (12).

Affinity-purified GST-NS5a fusion protein was cross-linked to an activated Affi-Gel 10 chromatography matrix (Bio-Rad). The affinity matrix thus obtained was then incubated for 1 h at 4°C with human sera and washed extensively with PBS, and the bound antibodies were eluted as described previously (12). In some preparations, anti-NS3 antibodies copurified with anti-NS5a immunoglobulins.

**In vitro transcription and translation.** Recombinant plasmids pCITE(146) and pSK(CORE) were linearized with restriction enzymes *Ssp*I and *Xho*I, respectively, and transcribed in vitro with T7 RNA polymerase as described previously (29). The transcripts were translated by using an mRNA-dependent rabbit reticulocyte lysate (Promega Biotec). All translation reactions were carried out in 50 μl in the presence or absence of canine pancreatic microsomal membranes (Promega Biotec). Translation mixtures were incubated at 30°C for 90 min with [$^{35}$S]methionine (Amersham) for labeling. Samples of translation mixtures were then immunoprecipitated with region-specific antisera, and the immunoprecipitated proteins were resolved by SDS-PAGE.

**Preparation of labeled cell extracts.** HeLa cells seeded at a density of $6 \times 10^5$ cells per plate were infected with vaccinia virus vTF7-3 at a multiplicity of 5 PFU per cell (19). After adsorption for 30 min at 37°C, 3 ml of Dulbecco's modified Eagle's MEM supplemented with 10% FCS was added. Cells were incubated an additional 30 min at 37°C. Twenty micrograms of recombinant plasmid DNA was precipitated in calcium phosphate as described previously (26) and added directly to each plate in a 500-μl volume. In the cotransfection experiments, 10 μg of each plasmid was precipitated in calcium phosphate. At 4 h posttransfection, the medium was replaced with MEM lacking methionine (GIBCO), and the cells were starved for 1 h at 37°C. Cells were then radiolabeled for 3 h with 400 μCi of $^{35}$S label (ICN) in 2 ml of MEM lacking methionine and supplemented with 2% dialyzed FCS. Cells were harvested and prepared for immunoprecipitation in $IPB_{150}$ (20 mM Tris-Cl [pH 8.0], 150 mM NaCl, 1% Triton) supplemented with 1 mM phenylmethylsulfonyl fluoride, 1 mM EDTA, and 1 mM dithiothreitol.

**Immunoprecipitation.** Prior to immunoprecipitation, SDS and dithiothreitol were added to the cell lysates to final concentrations of 2% and 10 mM, respectively. The lysates were then incubated at room temperature for 1 h and heated at 95°C for 10 min. Ten-microliter samples of antisera used in the immunoprecipitation reactions were preadsorbed for 1 h at 4°C in a 400-μl volume of $IPB_{150}$ with vT7F3-infected HeLa cell extracts spotted on nitrocellulose filters. The antibody suspension was then incubated with 60 μl of protein A (PA)-Sepharose for 1 h at 4°C. The PA-Sepharose beads were pelleted by centrifugation, washed three times in 1 ml of $IPB_{150}$, resuspended in 400 μl of $IPB_{150}$, and incubated for an additional hour at 4°C with 20 μl of cell lysate. All reactions were performed with constant mixing on an end-over-end rotator. The PA-Sepharose suspension was then layered on 0.9 ml of 0.5× NDET (0.5% Nonidet P-40, 0.2% sodium deoxycholate, 33 mM EDTA, 10 mM Tris-Cl [pH 7.4]) containing 30% sucrose and pelleted by centrifugation in a microcentrifuge for 10 min at room temperature. The pellet was resuspended in 300 μl of NDET–0.3% SDS and then washed twice with the same buffer and once with water (27). The sample was resuspended in 20 μl of sample buffer and heated at 95°C, and the supernatant was then subjected to SDS-PAGE.

TABLE 1. Region-specific antisera to E1, E2/NS1, NS2, NS3, NS4, and NS5

| Target protein | Predicted boundaries (amino acids)[a] | HCV cDNA fragment (nucleotides)[b] | Antiserum specificity |
|---|---|---|---|
| E1 | 193–383 | 990–1350 (220–340) | E1[c] |
| E2/NS1 | 384–729 | 1501–2525 (392–733) | E2/NS1 |
| NS2 | 731–1007 | 3018–3301 (896–990) | NS2 |
| NS3 | 1008–1616 | 3890–4716 (1187–1496) | NS3 |
| NS4 | 1617–2013 | 5184–6210 (1618–1960) | NS4a, NS4b |
| NS5 | 2014–3011 | 6940–7467 (2204–2379) | NS5a[c] |
| | | 8079–8926 (2585–2880) | NS5b |

[a] Boundaries of E1 and E2/NS1 on the HCV polyprotein were assigned as described by Hijikata et al. (14). The limits of the nonstructural proteins were assigned as described by Takamizawa et al. (31).

[b] cDNA fragments were cloned in pATH and pGEX-3x expression vectors as described in Materials and Methods. Amino acid positions are given in parentheses.

[c] Antibodies specific for this protein were purified from HCV-seropositive patients as described in Materials and Methods.

## RESULTS

**Production of fusion proteins and region-specific antisera.** To monitor the expression and processing of the HCV polyprotein, several region-specific antisera were prepared in rabbits against HCV fusion proteins expressed in bacteria. Human polyclonal antibodies specific for the N-terminal half of NS5 and for the E1 protein were purified from patient sera by immunoaffinity. Table 1 describes the cDNA fragments used for the construction of pATH/HCV and pGEX/HCV plasmids containing the relevant regions of the HCV genome. The cDNA fragments were chosen on the basis of putative boundaries of each HCV viral protein. These boundaries were established by comparing the HCV polyprotein sequence with that of flaviviruses and identifying the putative processing sites which could be responsible for the release of HCV proteins from the polyprotein precursor (14, 31).

**In vitro processing of the structural proteins of HCV.** Cell-free protein synthesis experiments were performed with truncated cDNA clones to examine the processing of HCV structural proteins. Figure 1 shows a diagram of all constructs used in this study. The results of in vitro processing assays using clones pCITE(146) and pSK(CORE) are shown in Fig. 2. The translation product of an RNA derived from plasmid pCITE(146) linearized at the *Ssp*I site at position 2873 was processed into three major proteins of 21, 37, and 61 kDa (Fig. 2, lane 1). The pattern of translation was significantly different when the reaction was carried out in the absence of microsomal membranes, as shown by the lack of the processed protein bands and by the presence of a large precursor which ran close to the origin of the gel (Fig. 2, lane 2). The 61-kDa protein was immunoprecipitated with a region-specific antiserum directed against the putative E2/NS1 region (Fig. 2, lane 3). The p21 polypeptide originates from the amino-terminal region of the polyprotein, and it was

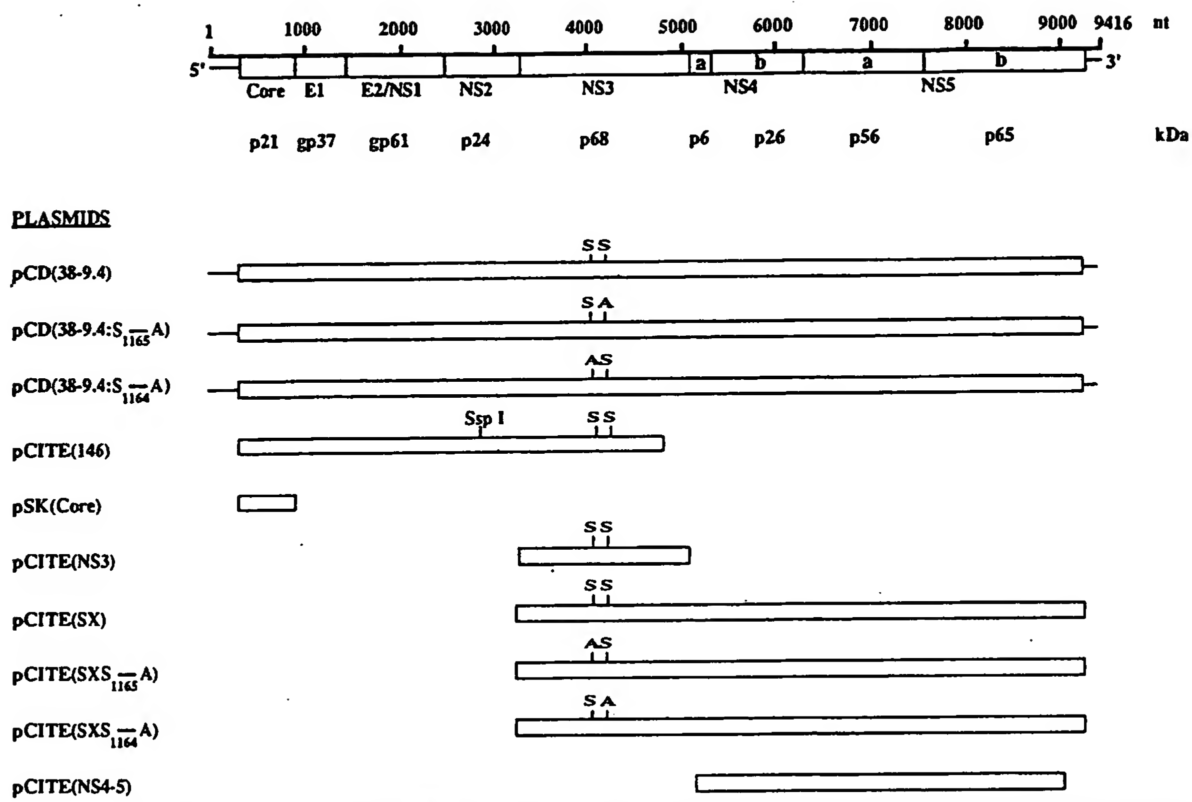



FIG. 1. Schematic representation of the HCV genome, predicted protein-coding domains, and recombinant expression plasmids used in this study. The viral genome is shown at the top (thin line, untranslated regions; open box, ORF); the molecular sizes of the viral proteins are indicated. The constructs expressing specific regions of the HCV genome are shown below. The name of each plasmid is shown on the left. The *Ssp*I site at nucleotide (nt) 2873 used to linearize plasmid pCITE(146) in the in vitro transcription-translation experiments is indicated. The letter designations outside the ORF box of the recombinant plasmids indicate the serine (S) residues at amino acid 1164 and 1165 which have been individually changed to an alanine (A) residue in several constructs.

identified as the core protein since it comigrated with the translation product derived from plasmid pSK(CORE) (Fig. 2, lane 4). The 37-kDa protein is derived from the middle portion of pCITE(146) and probably corresponds to the structural protein E1. Thus, the results are in agreement with the published data indicating that in vitro processing of the structural region of HCV polyprotein is dependent of the presence of microsomal membranes in the translation reaction, suggesting that it is mediated by cellular signal peptidases (14).

**Transient expression of HCV cDNA encoding the entire polyprotein.** Attempts to use cell-free protein synthesis to examine processing of the nonstructural region of HCV polyprotein were undermined by complex patterns of translation products, making it difficult to obtain conclusive results (data not shown). To examine the processing of the nonstructural region, we used the vaccinia virus T7 transient expression system (10). This system is based on the transfection of mammalian cells infected with a recombinant vaccinia virus expressing the bacteriophage T7 RNA polymerase. The T7 RNA polymerase produced by the recombinant vaccinia virus drives the expression of the transfected plasmid. Plasmid pCD(38-9.4) was constructed with the entire HCV ORF positioned downstream of bacteriophage T7 RNA polymerase promoter and was used for transfection experiments in HeLa cells. The transfected cells were labeled with [$^{35}$S]methionine, and the cell lysates were denatured in SDS and then immunoprecipitated with region-

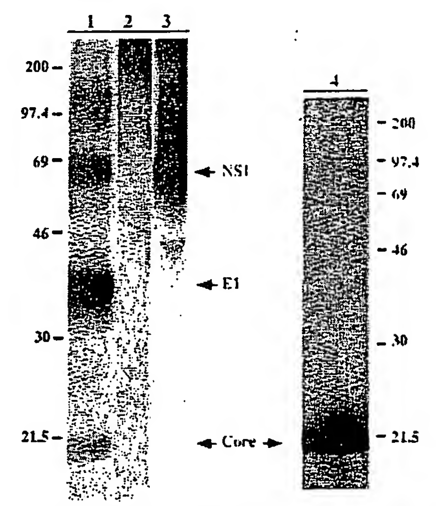



FIG. 2. Autoradiogram of SDS-PAGE analysis of the in vitro translation products. The transcripts of pCITE(146) (lanes 1 to 3) and of pSK(Core) (lane 4) were translated in vitro with a rabbit reticulocyte lysate in the presence (lanes 1, 3, and 4) or absence (lane 2) of microsomal membranes. Translation products labeled with [$^{35}$S]methionine were analyzed on an SDS–10% polyacrylamide gel directly (lanes 1, 2, and 4) or after immunoprecipitation with an anti-E2/NS1 antiserum (lane 3). Positions of the relevant translation products and of molecular weight standards (in kilodaltons) are indicated.

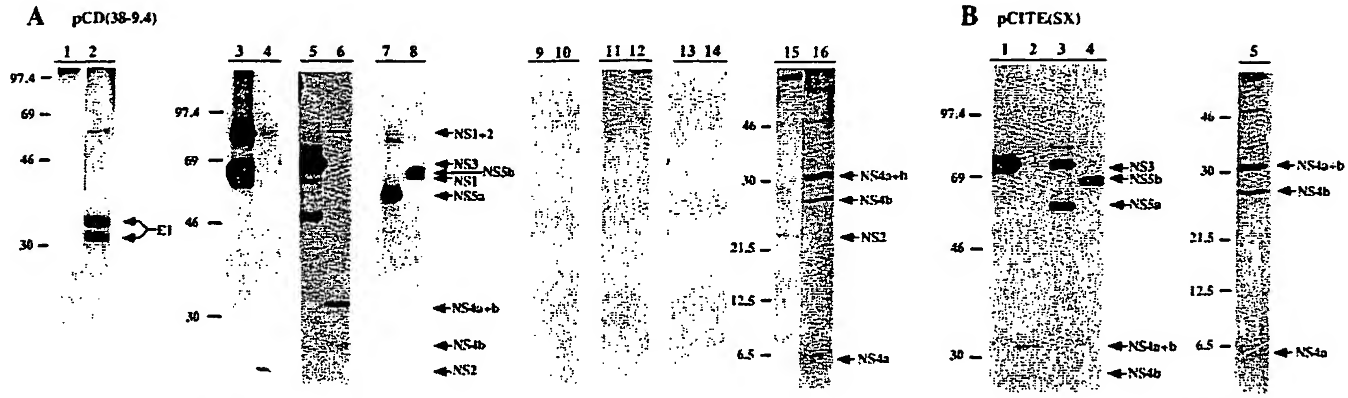



FIG. 3. Transient expression of plasmids pCD(38-9.4) and pCITE(SX) in mammalian cells. Recombinant plasmids pCD(38-9.4) and pCITE(SX) were transfected into vaccinia virus vT7F-3-infected HeLa cells as described in Materials and Methods. Cells were labeled with [$^{35}$S]methionine for 3 h, and SDS-denatured lysates were immunoprecipitated with antisera to specific regions of HCV ORF. Positions of the relevant HCV proteins immunoprecipitated with anti-HCV antibodies are indicated. Sizes of molecular weight standards are indicated in kilodaltons. (A) Lysates from cells transfected with pCD(38-9.4) were immunoprecipitated with anti-E1 (lane 2), anti-E2/NS1 (lane 3), anti-NS2 (lanes 4 and 15), anti-NS3 (lane 5), anti-NS4 (lanes 6 and 16), anti-NS5a (lane 7), and anti-NS5b (lane 8) antibodies. The immunoprecipitated products were resolved on an SDS–10% (lanes 1 to 14) or SDS–15% (lanes 15 and 16) polyacrylamide gel. Mock-transfected cell lysate was also immunoprecipitated with anti-HCV antibodies (lane 1, anti-E1; lane 9, anti-E2/NS1; lane 10, anti-NS2; lane 11, anti-NS3; lane 12, anti-NS4; lane 13, anti-NS5a; lane 14, anti-NS5b). (B) Lysates from cells transfected with pCITE(SX) were immunoprecipitated with anti-NS3 (lane 1), anti-NS4 (lanes 2 and 5), anti-NS5a (lane 3), and anti-NS5b antibodies (lane 4). Immunoprecipitated products were resolved on an SDS–10% (lanes 1 to 4) or SDS–15% (lane 5) polyacrylamide gel.

specific antisera. As shown in Fig. 3A, transfection of pCD(38-9.4) generated proteins that reacted specifically with anti-HCV antibodies. Two bands of 34 and 37 kDa were immunoprecipitated with the anti-E1 antiserum, indicating that these two polypeptides are derived from the E1 region (Fig. 3A, lane 2). The multiple bands of the expressed E1 protein may be due to incomplete processing or glycosylation in this experimental system. Anti-E2/NS1 antiserum immunoprecipitated a 61-kDa protein as observed in the in vitro translation studies; this protein presumably represents the mature form of E2/NS1. A second band of 77 kDa was recognized by anti-E2/NS1 and anti-NS2 antisera, suggesting that it may represent an uncleaved E2/NS1-NS2 precursor (Fig. 3A, lanes 3 and 4). A 24-kDa band was immunoprecipitated with the NS2-specific antiserum, consistent with the predicted molecular weight (Fig. 3A, lane 4). An NS3-related protein of 68 kDa was produced, as expected from the predicted molecular weight of this protein. A minor 47-kDa band was also consistently observed in the immunoprecipitation reactions, which could be due to an alternate cleavage of NS3 (Fig. 3A, lane 5). NS4-related proteins of 6, 26, and 31 kDa were immunoprecipitated with the region-specific antiserum, indicating that this region of the HCV polyprotein is cleaved into distinct protein products (Fig. 3A, lane 6). The NS2- and NS4-related proteins were clearly visible on an SDS–15% polyacrylamide gel (Fig. 3A, lanes 15 and 16). The 6-kDa protein was designated NS4a, and the 26-kDa protein species was called NS4b; the 31-kDa band probably represents the NS4a-NS4b precursor (see below).

Two proteins of 56 and 65 kDa were immunoprecipitated with the antisera specific for the N and C halves of the NS5 gene, respectively, indicating that the predicted 116-kDa protein is cleaved into at least two smaller products (Fig. 3A, lanes 7 and 8). The N-terminal protein was designated NS5a, whereas the C-terminal protein was called NS5b. In addition, the antisera specific for NS4 and NS5 recognized higher-molecular-weight bands which probably represent uncleaved precursors. These data suggest that apparently authentic processing of the entire HCV polyprotein can occur in this test system and that the function of NS3 as a viral protease can be determined.

In view of the observation that cleavage of the nonstructural region of the flavivirus polyprotein is dependent not only on the NS3 protease but also on the NS2b protein product, which may act as a cofactor for functional activity (3, 9), we determined whether the HCV polyprotein region encoding proteins NS3, NS4, and NS5 could be expressed and processed correctly despite the absence of the structural proteins and of NS2. To this end, a plasmid encompassing nucleotides 3303 to 9416 was constructed. This clone, named pCITE(SX), expresses all of NS3, NS4, and NS5 in addition to a few amino acid residues of NS2. As shown in Fig. 3B, transfection of pCITE(SX) generated proteins consistent in size with processed NS4a, NS4b, NS5a, and NS5b. The NS3 protein band immunoprecipitated with the anti-NS3 antiserum was slightly larger than that expressed by the full-length clone pCD(38-9.4), with an apparent molecular size of 74 kDa (Fig. 3B, lane 1). Furthermore, NS3 coprecipitated with the NS5a protein in several immunoprecipitation experiments because of the presence of anti-NS3 antibodies in some of the human anti-NS5a immunoglobulin preparations used in these studies (Fig. 3B, lane 3). These data indicate that processing of the N termini of NS4a, NS4b, NS5a, and NS5b is independent of the NS2 protein, while cleavage at the N terminus of NS3 may be affected by NS2 sequences.

**The NS3 catalytic domain is necessary for processing of the HCV polyprotein.** Site-directed mutagenesis was used to test the role of the putative NS3 protease domain in processing of the HCV polyprotein. Mutagenesis of this protein was accomplished by using synthetic oligonucleotides and PCR as described in Materials and Methods. Nucleotide 3825, T, was converted to a G, resulting in an amino acid change at position 1165. This mutation substitutes the catalytic serine residue with alanine. The mutant plasmid was designated pCD(38-9.4:$S_{1165}$-A). As a comparison, Ser-1164, which is not part of the putative catalytic triad, was changed to Ala by

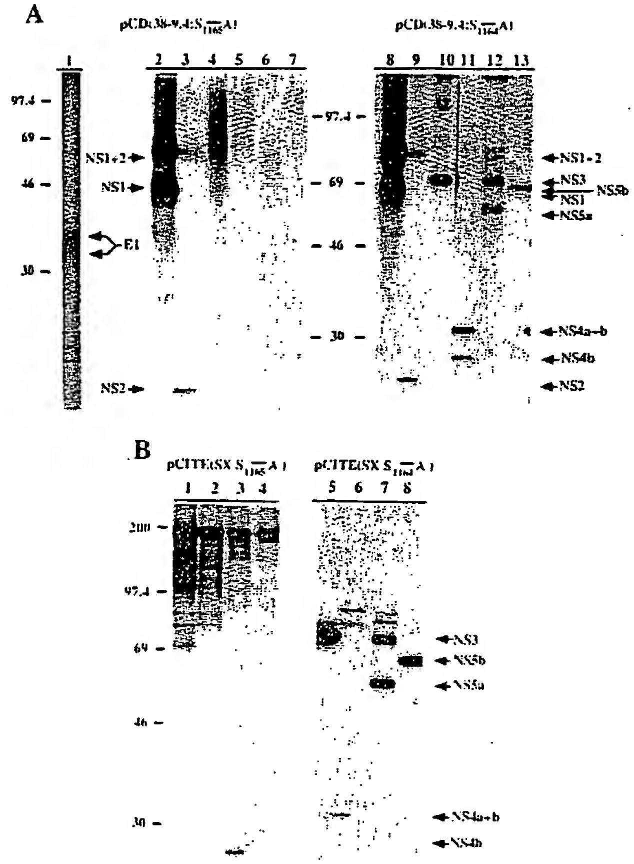



FIG. 4. Processing of the HCV ORF containing altered residues in NS3. Constructs containing a Ser-to-Ala substitution at the putative catalytic Ser-1165 or at Ser-1164 were transfected into vTF7-3-infected HeLa cells. Cell lysates were immunoprecipitated with anti-HCV antibodies, and HCV proteins were resolved on an SDS–10% polyacrylamide gel. Positions of relevant HCV proteins and of molecular weight standards (in kilodaltons) are indicated. (A) Lysates from cells transfected with plasmid pCD(38-9.4:$S_{1165}$-A) or pCD(38-9.4:$S_{1164}$-A) were immunoprecipitated with anti-E1 (lane 1), anti-E2/NS1 (lanes 2 and 8), anti-NS2 (lanes 3 and 9), anti-NS3 (lanes 4 and 10), anti-NS4 (lanes 5 and 11), anti-NS5a (lanes 6 and 12), and anti-NS5b (lanes 7 and 13) antibodies. (B) Lysates from cells transfected with plasmid pCITE(SX$S_{1165}$-A) or pCITE (SX$S_{1164}$-A) were immunoprecipitated with anti-NS3 (lanes 1 and 5), anti-NS4 (lanes 2 and 6), anti-NS5a (lanes 3 and 7), and anti-NS5b (lanes 4 and 8) antibodies.

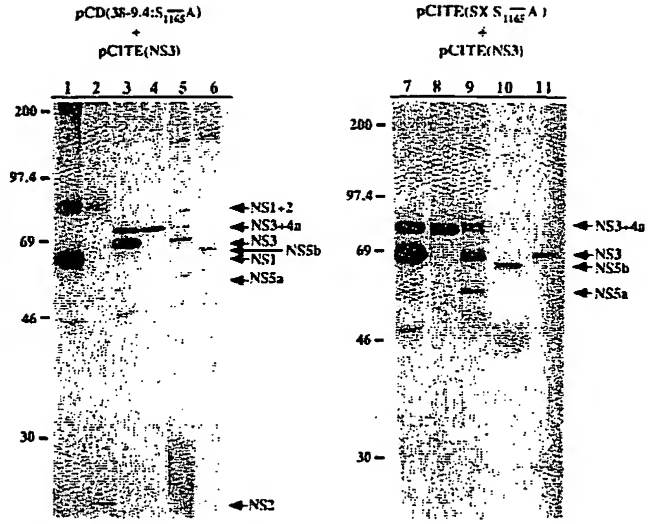



FIG. 5. Cotransfection experiments to examine *trans* cleavage of a catalytically inactive HCV ORF substrate. Plasmid pCITE(NS3) was cotransfected with pCD(38-9.4:$S_{1165}$-A) or pCITE(SX$S_{1165}$-A) as described in the legend to Fig. 3 and in Materials and Methods. Cell lysates were immunoprecipitated with anti-E2/NS1 (lane 1), anti-NS2 (lane 2), anti-NS3 (lanes 3 and 7), anti-NS4 (lanes 4 and 8), anti-NS5a (lanes 5 and 9), and anti-NS5b (lanes 6 and 10) antibodies and loaded on an SDS–10% polyacrylamide gel. NS3 protein immunoprecipitated from lysate of cells transfected with pCITE(NS3) alone is shown in lane 11. Positions of relevant HCV proteins and of molecular weight standards (in kilodaltons) are indicated.

substituting nucleotide 3822, T, with a G. This construct was designated pCD(38-9.4:$S_{1164}$-A). Mutagenesis was confirmed by sequencing the region of the mutation in both parent and mutant plasmids.

The transfection of plasmid pCD(38-9.4:$S_{1165}$-A) in HeLa cells resulted in the correct synthesis of the putative structural proteins, as indicated by the immunoprecipitation of E1 and E2/NS1 (Fig. 4A, lanes 1 and 2). NS2 was also released from the polyprotein precursor and immunoprecipitated with the region-specific antiserum (Fig. 4A, lane 3). The mature products of NS3, NS4a, NS4b, NS5a, and NS5b were not detectable in the transfected cell extracts, suggesting that the mutation of the putative catalytic Ser residue had specifically interfered with the cleavage of these proteins (Fig. 4A, lanes 4 to 7). A similar pattern of immunoprecipitations of uncleaved precursors was observed when the Ser-to-Ala substitution was introduced at position 1165 in construct pCITE(SX) (Fig. 4B, lanes 1 to 4). In contrast, the Ser-to-Ala substitution at residue 1164 did not have any effect on the processing profile of the HCV protein products, as shown by the transfection of plasmids pCD(38-9.4:$S_{1164}$-A) (Fig. 4A, lanes 8 to 13) and pCITE(SX$S_{1164}$-A) (Fig. 4B, lanes 5 to 8). Both the structural and nonstructural proteins were detected in their processed form. Thus, these results are consistent with the hypothesis that the NS3 protein is indeed a serine protease and that residue 1165 is the catalytic serine of this enzyme. Furthermore, they demonstrate that a functional NS3 protease domain is required for efficient cleavage at the NS3-NS4a, NS4a-NS4b, NS4b-NS5a, and NS5a-NS5b sites.

***trans*-cleavage activity of NS3.** The data presented above established that NS3 is required for the proper biogenesis of mature HCV proteins. To determine whether processing at the NS3-dependent sites could be mediated in *trans*, plasmid pCD(38-9.4:$S_{1165}$-A) was cotransfected with pCITE(NS3), a clone expressing amino acids 1007 to 1615 encompassing the entire NS3 protease domain. Figure 5 illustrates the results of the cotransfection experiments. As expected, coexpression of the NS3 protein with the entire polyprotein lacking a functional NS3 did not change the processing of the structural region and of NS2, since the cleavage of these proteins does not require NS3 (Fig. 5, lanes 1 and 2). The wild-type expression of NS5a and NS5b was restored, indicating that cotransfection of the NS3 clone with the mutated plasmid had abolished abnormal processing of these proteins (Fig. 5, lanes 5, 6, 10, and 11). In contrast, the NS4-specific antiserum immunoprecipitated a 78-kDa band not detected in cells transfected with pCD(38-9.4:$S_{1165}$-A) or pCITE (SX$S_{1165}$-A) alone. This protein, in addition to the NS3

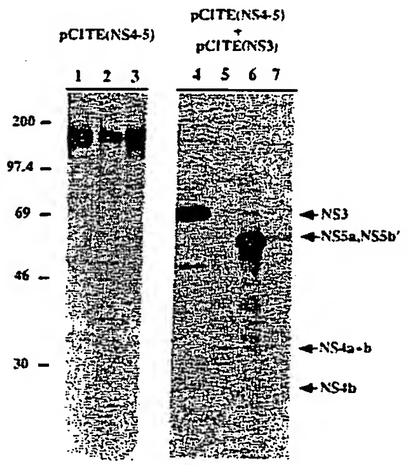



FIG. 6. Processing of HCV polyprotein substrate in *trans*. Plasmid pCITE(NS4-5) was transfected alone (lanes 1 to 3) or with clone pCITE(NS3) (lanes 4 to 7). Cell lysates were immunoprecipitated with anti-NS3 (lane 4), anti-NS4 (lanes 1 and 5), anti-NS5a (lanes 2 and 6), and anti-NS5b (lanes 3 and 7) antibodies. The truncated form of NS5b is indicated as NS5b'. Proteins were resolved on an SDS–10% polyacrylamide gel. Positions of molecular weight standards (in kilodaltons) and of relevant HCV products are indicated.

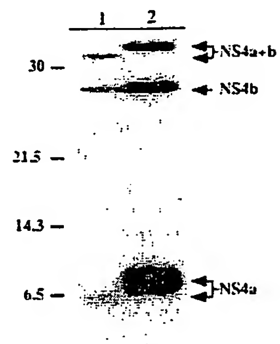



FIG. 7. Identification of NS4 precursor and mature products. Lysates from cells transfected with plasmid pCD(38-9.4) (lane 1) or pCITE(NS4-5) (lane 2) were immunoprecipitated with anti-NS4 antibodies. HCV proteins were resolved on an SDS–15% polyacrylamide gel. Positions of molecular weight standards (in kilodaltons) and of relevant HCV proteins are indicated.

product derived from pCITE(NS3), was recognized by the NS3-specific antiserum (Fig. 5, lanes 3, 4, 7, and 8). The 78-kDa protein probably represents the uncleaved NS3-NS4a precursor. The processed NS4a and NS4b products were not detected in the cotransfected cells, suggesting that expression of the NS4-related proteins had not been restored to normal by the cotransfection of pCITE(NS3) and pCD(38-9.4:$S_{1165}$-A) (Fig. 5, lanes 4 and 8).

To further examine processing of the NS4 and NS5 regions, cotransfection experiments were performed with pCITE(NS3) and pCITE(NS4-5). The latter plasmid contains nucleotides 5281 to 9071 encoding amino acids 1651 to 2921 which represent NS4 and part of NS5. Transfection of pCITE(NS4-5) alone resulted in the synthesis of uncleaved precursors which were immunoprecipitated by both NS4- and NS5-specific antisera (Fig. 6, lanes 1 to 3). In contrast, cotransfection with pCITE(NS3) resulted in the synthesis of mature forms of NS4a, NS4b, and NS5a and of the truncated form of NS5b, NS5b' (Fig. 6, lanes 4 to 7). Thus, cleavage at the N termini of NS4b, NS5a, and NS5b can be mediated in *trans* by NS3 whereas processing at the N terminus of NS4a cannot, suggesting that cleavage occurs as an intramolecular event (in *cis*).

To characterize in detail the NS4-related proteins, we compared the molecular weights of the NS4 polypeptides derived from the cotransfection of pCITE(NS3) and pCITE(NS4-5) with those of the NS4 products obtained from the transfection of pCD(38-9.4). As shown in Fig. 7, NS4a protein derived from pCD(38-9.4) is slightly smaller than that obtained from pCITE(NS4-5). This difference in molecular weight is probably due to the presence of NS3-related residues at the N terminus of NS4 derived from pCITE(NS4-5) which cannot be removed in *trans* by the HCV protease. This observation suggests that the actual border between NS3 and NS4 is located downstream of amino acid residue 1651. Furthermore, the observation that the difference in molecular weight is also observed for the putative NS4a+b precursor but not for the NS4b protein supports our interpretation of the 31-kDa polypeptide as the NS4a+b precursor; as observed for the NS4a polypeptide, the NS4a+b protein derived from plasmid pCITE(NS4-5) contains additional residues originating from NS3 and therefore is larger than the same protein derived from the wild-type plasmid.

## DISCUSSION

We have analyzed HCV protein biogenesis by in vitro translation studies and by transient expression in mammalian cells to characterize the sizes of the proteins expressed by specific regions of the HCV ORF and to determine the role of cellular protease and of the putative viral protease in HCV polyprotein processing. The results obtained confirm the previously reported observation that the structural region of HCV is processed by cellular proteases independently of the presence of the nonstructural proteins (14). Furthermore, they demonstrate that the NS3 protein is a serine protease and that a catalytically active NS3 is required for the proper cleavage of most of the nonstructural proteins.

A more detailed description of the genetic order and size of the proteins encoded by the HCV genome can now be provided on the basis of immunoprecipitation of virus-specific proteins with antisera directed against precise regions of the HCV ORF. In addition to the putative structural components of HCV, which have already been mapped to the N terminus of HCV ORF, other proteins can now be identified. Therefore, the HCV genetic order is the following: $NH_2$-p21-gp37-gp61-p24-p68-p6-p26-p56-p65-COOH. p21, gp37, and gp61 correspond to the structural proteins C, E1, and E2/NS1, respectively, as shown by the in vitro translation of the most N-terminal 850 amino acids (Fig. 2) and as reported by Hijikata et al. (14). Interestingly, the expression of E1 in transfected cells resulted in the synthesis of two proteins of 34 and 37 kDa which were both immunoprecipitated with an anti-E1 antiserum (Fig. 3). The pattern of expression of the E1 protein in transfected cells is therefore different from that observed in in vitro translation studies, in which a single protein product of 37 kDa is observed (Fig. 2). The two proteins corresponding to E1 expressed in transfected cells are probably due to incomplete glycosylation or processing of the viral glycoprotein

and have been observed both in mammalian and insect cells (19, 22).

The protein encoded by the NS2 region of HCV ORF is a polypeptide of 24 kDa. The size of this protein differs from that encoded by the NS2 gene of flaviviruses, indicating that no further processing of this region takes place other than the release of p24 from the precursor (2). Immunoprecipitation of a large 78-kDa protein with both anti-E2/NS1 and anti-NS2 antisera suggests that this protein represents an uncleaved E2/NS1-NS2 precursor. This observation is not unique to HCV but has been reported also for the processing of flaviviruses (2, 21), suggesting that the maturation of E2/NS1 and NS2 proteins is not a cotranslational event, but that it may occur at later stages of the biogenesis of these two polypeptides. Alternatively, expression of the HCV ORF in this particular system is higher than normal physiological levels such that the cleavage between E2/NS1 and NS2 may be a limiting event due to the shortage of specific cellular factors required for proper processing at this site.

NS3 antiserum recognizes a 68-kDa protein, in agreement with the predicted molecular weight of NS3. Interestingly, a minor 47-kDa band is also consistently recognized by anti-NS3 antibodies in immunoprecipitation experiments using denatured antigens (Fig. 3) as well as in Western immunoblots (data not shown). The 47-kDa protein therefore very probably does not represent a cellular product associated with the viral protease but rather is a further cleavage or degradation product derived from the NS3 region. The functional significance of this 47-kDa protein in HCV processing is unknown.

The NS4 region of the HCV ORF is cleaved into two distinct products, NS4a and NS4b, which are both recognized by the rabbit polyclonal antiserum. This observation is consistent with the processing profile of flavivirus nonstructural region (2). Although we do not have antisera directed against selected regions of NS4, it is very likely that the NS4a protein of 6 kDa is the most N-terminal part of the NS4 gene, whereas the NS4b protein of 24 kDa is encoded by the C-terminal region of the NS4 gene. This conclusion is in part supported by complementation studies of the mutated clones lacking a catalytically active NS3, which showed that the NS3-NS4 uncleaved product has the molecular weight which corresponds to the sum of the molecular weight of NS3 and NS4a (Fig. 5). Furthermore, we have consistently observed the presence of an NS4-related 31-kDa protein in our immunoprecipitation experiments, which we interpret as being the NS4a+b precursor.

The NS5 region of the HCV polyprotein is cleaved into two smaller products of 47 and 65 kDa; the processing of this region therefore differs from that of flavivirus NS5, which is released from the polyprotein precursor as a single protein of 110 kDa. The GDD consensus sequence characteristic of RNA-dependent RNA polymerases is located in NS5b (residues 2736 to 2738), indicating that this protein may act as a viral RNA replicase during HCV-specific RNA synthesis (17). However, NS5a could also have a function in the replication of the viral genome, acting as a component of the replication complex involved in the reaction.

Processing of C, E1, and E2/NS1 is mediated by signal peptidases located in the endosplasmic reticulum lumen of the host cell as in the case of the flavivirus structural proteins. This conclusion is based on in vitro translation studies, in which mature forms of the putative structural components of HCV are observed when the translation reaction is carried out in the presence of microsomal membranes. A similar result has been obtained by Hijikata et al. (14), who have indicated the presence of hydrophobic segments located at the N termini of E1 and E2/NS1 (residues 174 to 191 and 371 to 383, respectively) which could act as signal sequences. Furthermore, transfection experiments with plasmid pCD(38-9.4:$S_{1165}$-A), in which the catalytic Ser residue has been substituted with Ala, indicates that the expression and processing of the E1 and E2/NS1 proteins are not affected by this mutation, suggesting that the processing of the structural region is independent of the NS3 protease. Similarly, cleavage at the N termini of NS2 and NS3 is also not dependent on the NS3 protease, since the NS2 protein is properly processed in this plasmid (Fig. 4). The nature of the protease(s) responsible for cleavage at these two sites is unknown; however, release of the flavivirus NS2a protein from the polyprotein precursor is thought to be mediated by a cellular protease, and a similar mechanism may generate the NS2 protein of HCV (8, 15). Interestingly, there seems to be a requirement for NS2 in the processing of the N terminus of NS3, as shown by the transfection of pCITE(SX) (Fig. 3B). In this construct, most of NS2 has been removed, with a concomitant impairment of the cleavage at the NS2-NS3 site demonstrated by the larger size of the viral protease (Fig. 3B). In view of the hydrophobic nature of the NS2 polypeptide, it is possible that the NS2 protein is required for correct localization of the NS2-NS3 protein in the cytoplasm, which renders the precursor available to cellular protease(s) for proper cleavage. Recently, the cellular localization of the dengue type 2 virus NS3-NS4a-NS4b-NS5 precursor protein was shown to be distinctly different from the perinuclear localization of the mature NS5 protein, and the distribution of these proteins within the cytoplasm has been suggested to be determined, at least in part, by the NS2b polypeptide (35).

The data reported here indicate that the NS3 protease activity is required for processing of the nonstructural proteins NS3, NS4, and NS5. The NS3 protease acts in *cis* at its C terminus to release itself from the polyprotein precursor in a fashion similar to that of the viral protease of flaviviruses and pestiviruses. This conclusion is based on the observation that cleavage at the N terminus of NS4a cannot be complemented in *trans*, whereas cleavage activity at the N termini of NS4b, NS5a, and NS5b can be demonstrated in *trans* (Fig. 5 and 6). The N termini of NS5a and NS5b are processed when pCITE(NS3) is cotransfected with the full-length clone pCD(38-9.4:$S_{1165}$-A) or with clones pCITE ($SXS_{1165}$-A) and pCITE(NS4-5), whereas the processed NS4b product is detected only when pCITE(NS3) is cotransfected with plasmid pCITE(NS4-5) (Fig. 6). A possible explanation of the difference in protein profile in these cotransfection experiments may be that detection of the NS4b protein in transfected cells depends on the presence of processed NS4a. Possibly, NS4a and NS4b form a complex which determines the stability of both polypeptides in transfected cells. However, the lack of release of NS4a from the polyprotein precursor may undermine this interaction, compromising the stability of NS4b and therefore preventing its detection in the transfected cells. Alternatively, the cleavage at the N terminus of NS4b may not be complemented in constructs pCD(38-9.4:$S_{1165}$-A) and pCITE($SXS_{1165}$-A), resulting in the lack of detection of the processed NS4b. This latter possibility is less probable since an uncleaved precursor with a molecular size corresponding to the sum of the sizes of NS3, NS4a, and NS4b (approximately 100 kDa) is not detected in the transfected cells (Fig. 5).

The data presented here are in complete agreement with sequence alignment studies which had predicted that NS3 is

a serine protease important for HCV polyprotein processing and therefore assigned a pivotal role to this protein in the biogenesis of mature HCV polypeptides. The exact location of the protease domain on the amino acid sequence of NS3 is not yet available but is predicted to be near the N terminus of this protein (23). Although no deletion studies have been presented, the identification of Ser-1165 as the catalytic residue clearly shows that the catalytic site is located close to the putative N terminus of NS3.

Although the N terminus of each protein has been roughly positioned on the HCV polyprotein on the basis of similarity to flaviviruses, no information is available concerning the exact sequence at the cleavage sites with the exception of the data provided by Hijikata et al., who have identified the N termini of E1 and E2/NS1 (14). Efforts directed toward obtaining direct amino-terminal sequence data should be highly rewarding because such data will allow a better definition of HCV polyprotein map and provide useful information on the sequence requirement for the cleavage activity of NS3.

The results described here provide an important description of the genetic order of the HCV proteins on the viral polyprotein and of the processing events required for the biogenesis of the HCV polypeptides. This information reinforces the genetic similarity between HCV and flaviviruses and pestiviruses, substantiating the common genetic origin which places these viruses in the same family. It is clear, however, that the results obtained with this transient expression system may not faithfully reproduce the proteolytic events which take place during HCV infection. It is possible that the level of protein expression obtained in this system may be much higher than normal, affecting important equilibria between precursors and proteases, which in turn may regulate HCV replication and protein synthesis. It is also impossible at this time to correlate the proteolytic activity of NS3 with virus replication. This type of consideration awaits the development of an in vitro infection system or of an infectious cDNA clone with which it should be possible to examine more closely the intracellular events that regulate HCV replication and protein biogenesis.

## ACKNOWLEDGMENTS

We thank H. Okayama (Osaka University) for supplying HCV cDNA subclones, G. Paonessa for vaccinia virus vTF7-3, P. Costa for rabbit immunizations, P. Neuner for oligonucleotide synthesis, Y. Cully for graphics, R. Cortese and J. Jiricny for critical review, and IRBM coworkers for helpful discussions.

## REFERENCES

1. **Cahour, A., B. Folgout, and C. J. Lai.** 1992. Cleavage of the dengue virus polyprotein at the NS3/NS4a and NS4b/NS5 junctions is mediated by viral protease NS2b-NS3, whereas NS4a/NS4b may be processed by a cellular protease. J. Virol. 66:1535–1542.
2. **Chambers, T. J., D. W. McCourt, and C. M. Rice.** 1990. Production of yellow fever virus proteins in infected cell: identification of discrete polyprotein species and analysis of cleavage kinetics using region-specific antisera. Virology 177: 159–174.
3. **Chambers, T. J., R. C. Weir, A. Grakoui, D. W. McCourt, F. F. Bazan, R. J. Fletterick, and C. M. Rice.** 1990. Evidence that the N-terminal domain of nonstructural protein NS3 from yellow fever virus is a serine protease responsible for site-specific cleavages in the viral polyprotein. Proc. Natl. Acad. Sci. USA 87:8898–8902.
4. **Choo, Q.-L., G. Kuo, A. J. Weiner, L. R. D. W. Bradley, and M. Houghton.** 1989. Isolation of a cDNA clone derived from a blood-borne non-A non-B viral hepatitis genome. Science 244: 359–362.
5. **Choo, Q.-L., K. H. Richman, J. H. Han, K. Berger, C. Lee, C. Dong, C. Gallegos, D. Coit, A. Medina-Selby, P. J. Barr, A. J. Weiner, D. W. Bradley, G. Kuo, and M. Houghton.** 1991. Genetic organization and diversity of the hepatitis C virus. Proc. Natl. Acad. Sci. USA 88:2451–2455.
6. **Collet, M. S., D. K. Anderson, and E. Retzel.** 1988. Comparison of the pestivirus bovine diarrhoea virus with members of the Flaviviridae. J. Gen. Virol. 69:2637–2643.
7. **Collet, M. S., V. Moennig, and M. C. Horzineck.** 1989. Recent advances in pestivirus research. J. Gen. Virol. 70:253–266.
8. **Falgout, B., R. Chanock, and C. J. Lai.** 1989. Proper processing of dengue virus nonstructural glycoprotein NS1 requires N-terminal hydrophobic signal sequence and the downstream nonstructural protein NS2a. J. Virol. 63:1852–1860.
9. **Falgout, B., M. Pethel, Y.-M. Zhang, and C.-J. Lai.** 1991. Both nonstructural proteins NS2b and NS3 are required for the proteolytic processing of dengue virus nonstructural proteins. J. Virol. 65:2467–2475.
10. **Fuerst, T. R., E. G. Niles, F. W. Studier, and B. Moss.** 1986. Eukaryotic transient-expression system based on recombinant vaccinia virus that synthesizes bacteriophage T7 RNA polymerase. Proc. Natl. Acad. Sci. USA 83:8122–8126.
11. **Harada, S., Y. Watanabe, K. Takeuchi, T. Suzuki, T. Katayama, Y. Takebe, I. Saito, and T. Miyamura.** 1991. Expression of processed core protein of hepatitis C virus in mammalian cells. J. Virol. 65:3015–3021.
12. **Harlow, E., and D. Lane.** 1988. Antibodies: a laboratory manual. Cold Spring Harbor Laboratory Press, Cold Spring Harbor, N.Y.
13. **Higuchi, R., B. Krummel, and R. S. Saiki.** 1988. A general method of *in vitro* preparation and specific mutagenesis of DNA fragments: study of protein and DNA interaction. Nucleic Acids Res. 16:7351–7367.
14. **Hijikata, M., N. Kato, Y. Ootusyama, M. Nakagawa, and K. Shimotohno.** 1991. Gene mapping of the putative structural region of the hepatitis C virus genome by in vitro processing analysis. Proc. Natl. Acad. Sci. USA 88:5547–5551.
15. **Hori, H., and C. J. Lai.** 1990. Cleavage of dengue virus NS1-NS2a requires an octapeptide at the C terminus of NS1. J. Virol. 64:4573–4577.
16. **Houghton, M., A. Weiner, J. Han, G. Kuo, and Q.-L. Choo.** 1991. Molecular biology of the hepatitis C viruses: implication for diagnosis, development and control of viral disease. Hepatology 14:381–388.
17. **Kamer, G., and P. Argos.** 1984. Primary structural comparison of RNA-dependent polymerases from plant, animal, and bacterial viruses. Nucleic Acids Res. 12:7269–7282.
18. **Kato, M., M. Hijikata, Y. Ootsuyama, M. Nakagawa, S. Ohkoshi, T. Sugimura, and K. Shimotohno.** 1990. Molecular cloning of human hepatitis C virus genome from Japanese patients with non-A non-B hepatitis. Proc. Natl. Acad. Sci. USA 87:9524–9528.
19. **Kohara, M., K. Tsukiyama-Kohara, N. Maki, K. Asano, K. Yamaguchi, K. Miki, S. Tanaka, N. Hattori, Y. Matsuura, I. Saito, T. Myamura, and A. Nomoto.** 1992. Expression and characterization of glycoprotein gp35 of hepatitis C virus using recombinant vaccinia virus. J. Gen. Virol. 73:2313–2318.
20. **Kuo, G., Q.-L. Choo, H. J. Alter, G. L. Gitnick, A. G. Redecker, R. H. Purcell, T. Myamura, J. L. Dienstag, M. J. Alter, C. E. Syevens, G. E. Tagtmeier, F. Bonino, M. Colombo, W.-S. Lee, C. Kuo, K. Berger, J. R. Shister, L. R. Overby, D. W. Bradley, and M. Houghton.** 1989. An assay for circulating antibodies to a major etiologic virus of human non-A non-B hepatitis. Science 244:362–364.
21. **Mason, P. W.** 1989. Maturation of Japanese encephalitis virus glycoproteins produced by infected mammalian and mosquito cells. Virology 169:354–364.
22. **Matuura, Y., S. Harada, R. Suzuki, Y. Watanabe, Y. Inoue, I. Saito, and T. Myamura.** 1992. Expression of processed envelope protein of hepatitis C virus in mammalian and insect cells. J.

Virol. 66:1425–1431.
23. Miller, R. H., and R. H. Purcell. 1990. Hepatitis C virus shares amino acid sequence similarity with pestiviruses and flaviviruses as well as members of two plant virus subgroups. Proc. Natl. Acad. Sci. USA 87:2057–2061.
24. Rice, C. M., E. M. Lenches, S. R. Eddy, S. R. Shin, R. L. Sheets, and J. H. Strauss. 1985. Nucleotide sequence of yellow fever virus: implication of flavivirus gene expression and evolution. Science 229:726–733.
25. Rice, C. M., E. G. Strauss, and J. H. Strauss. 1986. Structure of the flavivirus genome, p. 279–326. *In* S. Schlesinger and M. J. Schlesinger (ed.), Togaviridae and Flaviviridae. Plenum Press, New York.
26. Sambrook, J., E. F. Fritsch, and T. Maniatis. 1989. Molecular cloning: a laboratory manual, 2nd ed. Cold Spring Harbor Laboratory Press, Cold Spring Harbor, N.Y.
27. Shin, S. U., and S. L. Morrison. 1989. Production and properties of chimeric antibody molecules. Methods Enzymol. 178:459–476.
28. Smith, D. B., and K. S. Johnson. 1988. Single-step purification of polypeptides expressed in *Escherichia coli* as fusions with glutathione S-transferase. Gene 67:31–40.
29. Soe, L. H., S.-K. Shieh, S. C. Baker, M.-F. Chang, and M. M. C. Lai. 1987. Sequence and translation of the murine coronavirus 5'-end genomic RNA reveals the N-terminal structure of the putative RNA polymerase. J. Virol. 61:3968–3976.
30. Spindler, K. R., S. E. Rosser, and A. J. Berk. 1984. Analysis of adenovirus transforming proteins from early regions 1A and 1B with antisera to inducible fusion antigens in *Escherichia coli*. J. Virol. 49:132–141.
31. Takamizawa, A., C. Mori, I. Fuke, S. Manabe, S. Murakami, J. Fujita, E. Onoshi, T. Andoh, I. Yoshida, and H. Okayama. 1991. Structure and organization of the hepatitis C virus genome isolated from human carriers. J. Virol. 65:1105–1113.
32. Takeuchi, K., Y. Kubo, S. Boonmar, Y. Watanabe, T. Kakayama, Q.-L. Choo, G. Kuo, M. Houghton, I. Saito, and T. Myamura. 1990. The putative nucleocapsid and envelope protein genes of hepatitis C virus determined by comparison of the nucleotide sequence of two isolates derived from an experimentally infected chimpanzee and healthy human carriers. J. Gen. Virol. 71:3027–3033.
33. Weiland, E., R. Stark, B. Haas, T. Rumenopf, G. Meyers, and H. J. Thiel. 1990. Pestivirus glycoprotein which induces neutralizing antibodies forms part of a disulfide-linked heterodimer. J. Gen. Virol. 64:3563–3569.
34. Wiskerchen, M., and M. S. Collet. 1991. Pestiviruses gene expression: protein p80 of bovine viral diarrhoea virus is a proteinase involved in polyprotein processing. Virology 184: 341–350.
35. Zhang, L., P. M. Mohan, and R. Padmanabhan. 1992. Processing and localization of dengue virus type 2 polyprotein precursor NS3-NS4a-NS4b-NS5. J. Virol. 66:7549–7554.

# Detection of Antigens Related to Hepatitis C Virus RNA Encoding the NS5 Region in the Livers of Patients with Chronic Type C Hepatitis

MIKIHIRO TSUTSUMI,[1] SACHIO URASHIMA,[1] AKIRA TAKADA,[1] TAKAYASU DATE[2] AND YUJIRO TANAKA[3]

[1]*Division of Gastroenterology, Department of Internal Medicine,* [2]*Department of Biochemistry, Kanazawa Medical University, Uchinada, Ishikawa 920-02, and* [3]*Second Department of Medicine, Tokyo Medical and Dental University, Tokyo 113, Japan*

**Hepatitis C virus is a positive single-strand RNA virus distantly related to flaviviruses. Therefore RNA replicase, an RNA-dependent RNA polymerase, may be essential for the replication of hepatitis C virus, as well as other RNA viruses. In this study we synthesized the recombinant polypeptide (HCV-NS5 antigen) with a 576 bp cDNA encoding a part of the NS5 region of the HCV genome that has the Gly-Asp-Asp motif. The antibody against this polypeptide was obtained from rabbit serum. In Western-blot analysis with NS5 IgG HCV antibody, an 84-kD protein was clearly detected as a single band in the microsomal fraction but not in the nuclear and mitochondrial fractions or in the cytosol fraction. Immunohistochemically, HCV-NS5 antigen was clearly stained in the cytoplasm of hepatocytes but not in the nucleus or cell membrane. Moreover, as determined on immunoelectron microscopy, HCV-NS5 antigen was demonstrated with fine granular distribution along the endoplasmic reticulum but not in other organelles, including the nucleus and mitochondria. Immunoreaction in other cell types was negative. These results indicate that replication of HCV may occur only in hepatocytes and that HCV-NS5 may be produced in the endoplasmic reticulum of these cells. HCV-NS5 antigen was stained only in the livers of hepatitis C virus–positive patients but not in sections from patients with chronic type B hepatitis or alcoholic fibrosis. In chronic type C liver disease, the overall detection rate of HCV-NS5 antigen was 56% (33% in chronic persistent hepatitis, 52% in chronic active hepatitis and 86% in cirrhosis). These results indicate that the replication of HCV may occur more frequently in the advanced than in the early stages of type C hepatitis. Recently, interferon has been used for treatment of chronic type C hepatitis. However, it is very difficult to determine when HCV is eliminated from the liver, even if hepatitis C virus RNA is not detected in serum. Immunostaining of HCV-NS5 antigen in liver biopsy sections may be helpful in evaluating the cessation of HCV replication. (HEPATOLOGY 1994;19:265-272.)**

Received March 26, 1993; accepted September 7, 1993.

Address reprint requests to: Akira Takada, M.D., Division of Gastroenterology, Department of Internal Medicine, Kanazawa Medical University, Uchinada, Ishikawa, 920-02, Japan.



0270-9139/94 $1.00 + .10 31/1/51540

Hepatitis C virus (HCV) is a major causative agent of non-A, non-B hepatitis that is associated with the development of cirrhosis and HCC (1). Although the mechanism of liver injury by HCV infection is still unknown, the replication of HCV does play a key role in the development of liver injury. However, the replication of HCV demands the participation of a specific enzyme capable of forming new RNA strands on parenteral RNA templates, because HCV is a plus-strand RNA virus (1, 2). An RNA-dependent RNA polymerase, called *RNA replicase,* may therefore be necessary for the replication of HCV.

Recently, we reported a method for the detection of HCV RNA genomes encoding a part of the NS5 region in patients' plasma (3, 4). The 401 bp cDNA encodes the Gly-Asp-Asp motif of RNA replicase, an important position for its activity (5). In this study we synthesized a recombinant polypeptide (HCV-NS5 antigen) with a cDNA encoding the Gly-Asp-Asp motif, and using the rabbit antibody against this polypeptide, we determined the intralobular distribution of HCV-NS5 antigen, which may include RNA-dependent RNA polymerase (HCV polymerase), immunohistochemically. Immunoelectron microscopic investigation was also used to clarify the intracellular site of replication of HCV.

## MATERIALS AND METHODS

***Patients.*** Needle-biopsy samples of human liver were obtained from 34 patients with chronic type C liver disease, including 6 with chronic persistent hepatitis (CPH), 21 with CAH, 7 with cirrhosis, 5 with alcoholic fibrosis and 5 with type B chronic hepatitis (Table 1). A part of each liver specimen (about 60% to 65%) was used for diagnostic purposes. The remainder was embedded in Tissue-Tek medium (Miles Inc., Diagnostics Div., Tarrytown, NY) after fixation with 2% periodate-lysine paraformaldehyde and was stored at −20° C. HCV antibody (anti-HCV) was determined in serum with a second-generation enzyme immunoassay (Dainabot Co., Ltd., Tokyo, Japan) (6), and HCV RNA genome encoding the NS-5 region (HCV-NS5) was detected with the one-stage reverse-transcription polymerase chain reaction (RT-PCR) method as reported previously (3, 4). In the HCV-NS5 negative serum,

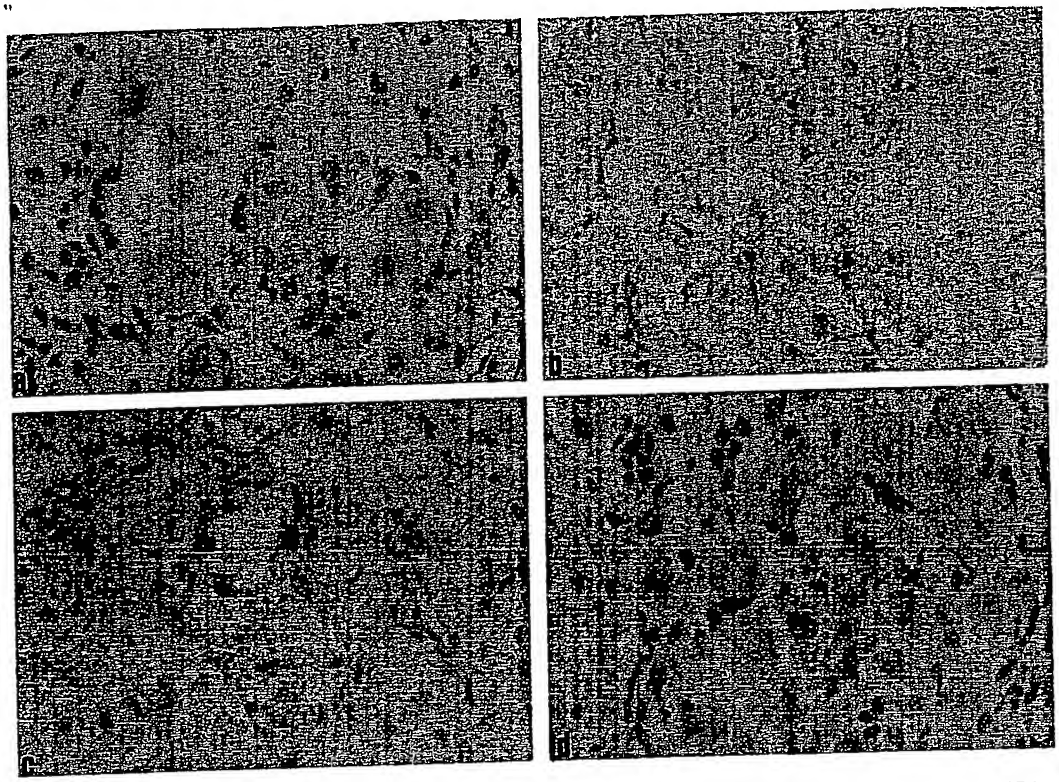


FIG. 1. Immunostaining of HCV-NS5 antigen in liver sections with chronic type C hepatitis (×400). (a) HCV-NS5 antigen stained in hepatocytes with anti-HCV-NS5 IgG. (b) Negative immunochemical reaction in the same section when preimmune rabbit IgG is substituted for HCV-NS5 IgG. (c) Prior absorption of anti-HCV-NS5 IgG with its immunogen completely abolishes specific immunostaining. (d) Positive immunostaining of HCV-NS5 antigen is not abolished by pretreatment with RNase.

TABLE 1. Subjects and detection rate of HCV-NS5 by immunostaining and HCV RNA by *in situ* hybridization in liver sections

| HCV marker status | HCV-NS5 antigen (%) | HCV RNA (%) |
|---|---|---|
| HCV marker–positive | 19/34 (55.9) | 23/27[a] (85.2) |
| CPH | 2/6 (33.3) | 2/3 (66.7) |
| CAH | 11/21 (52.4) | 18/21[a] (85.7) |
| Cirrhosis | 6/7 (85.7) | 3/3 (100) |
| HCV marker–negative | 0/10 (0) | 0/10 (0) |
| Chronic type B hepatitis | 0/5 (0) | 0/5 (0) |
| Alcoholic fibrosis | 0/5 (0) | 0/5 (0) |

[a] $p < 0.02$ vs. HCV-NS5 antigen by the $\chi^2$ test.

HCV RNA genome of the 5′-noncoding region (HCV-5′NC) was also detected with the two-stage PCR method as described by Okamoto et al. (7). HBsAg and HBV core antibody (anti-HBc) were determined as markers of HBV infection. In patients found to have type C chronic liver disease, at least one of the serum HCV markers was positive, but HBV markers were negative. In patients found to have chronic type B liver disease, HBs-antigen was positive, but all HCV markers were negative. In patients with alcoholic fibrosis, both HCV and HBV markers were negative (Table 1). Liver specimens were also obtained from seven patients with chronic type C hepatitis immediately after treatment with interferon for 6 mo. In all of these patients, HCV-NS5 antigen in a liver section was positive before interferon treatment.

***Preparation of Antibody.*** A cDNA encoding a portion of the HCV-NS5 region between residues 2,636 and 2,827 was obtained by means of RT-PCR from HCV-K1 (4) and was inserted into pET8c vector carrying a promoter for T7-RNA polymerase (8). Nucleotide sequence analysis indicated that the sequence contained nucleotides without stop codon and included Gly-Asp-Asp motifs near the center of the amplified fragment (residues 2,736-2,738). Peptides of 26 kD and 23 kD were synthesized in *Escherichia coli* and comprised as much as 90% of total bacterial proteins. Sequential Edman degradation analysis of the $NH_2$-terminal sequence of the two products indicated that they were derived from HCV cDNA and that 23-kD products were generated from a 26-kD polypeptide by processing of the C-terminal domain. Antibody was then raised against these purified polypeptides (HCV-NS5 antigen) in male New Zealand white rabbits with standard immunization techniques. The IgG fraction (anti-HCV-NS5 IgG) was purified by means of caprylic acid–ammonium sulfate fractionation (9).

***Immunohistochemical Procedure.*** Immunohistochemical

and immunoelectron microscopic studies were performed in blinded fashion to evaluate the results of virus markers in blood. Frozen liver sections 5 to 6 μm thick were cut from embedded tissue blocks on a cryostat at −20° C and air dried. After rehydration in PBS (100 mmol/L, pH 7.4), tissue sections were initially treated for 5 min with 0.03% hydrogen peroxide in methanol to eliminate endogenous peroxidase activity. After washing with PBS, sections were incubated with normal goat serum for 5 min to block nonspecific binding protein. Sections were additionally incubated with blood group antigens A, B and H (Dako Japan Co., Ltd., Kyoto, Japan) for 30 min at room temperature to prevent nonspecific reactions. After treatment with avidin-blocking solution (Dako Japan Co., Ltd.) and biotin-blocking solution (Dako Japan Co., Ltd.) for 15 min each, sections were incubated with anti-HCV-NS5 IgG for 30 min at room temperature. After washing with PBS, sections were incubated with biotin-labeled goat anti-rabbit IgG for 10 min at room temperature. Sections were then incubated with peroxidase-labeled streptavidin for 10 min at room temperature. After washing with PBS, peroxidase activity was developed in 50 mmol/L Tris-HCl buffer, pH 7.4, containing 3,3′-diaminobenzidine (0.3 mg/ml) and hydrogen peroxide (0.05%) for 5 min. Sections were subsequently counterstained with methylgreen or hematoxylin. Three types of control reactions were performed: (a) substitution of an equivalent amount of preimmune rabbit IgG for the primary antibody (rabbit anti-HCV NS5 IgG), (b) use of primary antibody that had been preabsorbed with its antigen and (c) omission of the primary antibody.

For the immunoelectron microscopic study, sections treated with anti-HCV-NS5 IgG were fixed with 1% osmium tetroxide in a veronal acetate buffer after development of peroxidase activity with diaminobenzidine. After dehydration with graded alcohols, the sections were embedded in Epon 812 (Serva Feinbiochemica GMBH & Co., Heidelberg, Germany). Ultrathin sections were obtained with the use of a diamond knife on an LKB-4800 ultramicrotome (LKB-Produkter AB, Bromma, Sweden), stained with lead hydroxide and examined under an electron microscope (Hitachi H-500; Hitachi Denshi Ltd., Tokyo, Japan).

*Other Methods.* Liver biopsy specimens, which were positive for staining of HCV-NS5 antigen, were homogenized in ice-cold 50 mmol/L Tris-HCl buffer, pH 7.4, containing 1.15% KCl, 1 mmol/L EDTA, 0.5 mmol/L phenylmethylsulfonyl fluoride and 0.02 mmol/L butylated hydroxytoluene (homogenate fraction) and were then centrifuged at 9,000 *g* for 30 min. Nuclear and mitochondrial fractions were obtained from the pellet. After calcium chloride was added to a final concentration of 8 mmol/L, the supernatant was recentrifuged at 15,000 *g* for 30 min. A cytosol fraction was obtained from the supernatant, and a microsomal fraction was obtained from the pellet (10).

SDS-PAGE of the liver fractions was performed according to the method of Laemmli (11) with a separating gel containing 7.5% acrylamide. After electrophoresis, the resolved proteins were transferred to nitrocellulose as described by Nielsen et al. (12). The protein blots were immunochemically stained by means of consecutive incubation with rabbit anti-HCV-NS5 IgG and horseradish peroxidase–conjugated goat antirabbit IgG. Blocking, washing and antibody incubation steps were all performed at room temperature in PBS containing 5% (wt/vol) nonfat dry milk protein. Peroxidase activity was subsequently localized on the nitrocellulose filter with 4-chloro-naphthol (2.8 mmol/L in 17% methanol) as chromogen (13).

A portion of the liver specimens obtained from many of the patients was also embedded in Tissue-Tek medium after fixation with 4% paraformaldehyde and was stored at −20° C. On the frozen liver sections, HCV RNA was stained by means of *in situ* hybridization with a cDNA probe for the core region according to the method of Tanaka et al. (14).

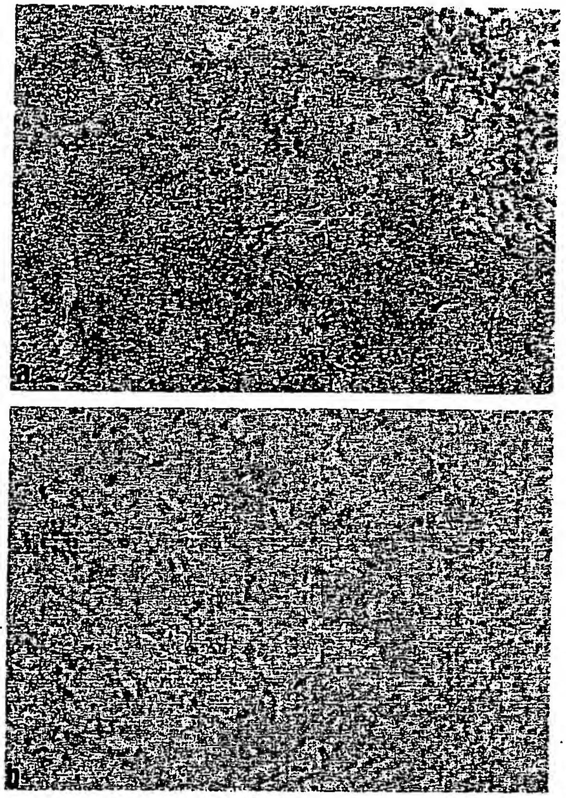


FIG. 2. Staining reactions of HCV RNA by *in situ* hybridization in liver sections with chronic type C hepatitis (×200). (a) HCV RNA stained only in hepatocytes. (b) Positive staining of HCV RNA is abolished by pretreatment with RNase.

## RESULTS

HCV-NS5 antigen was clearly stained in hepatocytes of patients with chronic type C liver disease (Fig. 1a). The immunostaining of HCV-NS5 antigen was observed only in the cytoplasm of hepatocytes in a granular fashion and not in the nucleus or cell membrane. In other types of liver cells, including infiltrating lymphocytes, HCV-NS5 antigen was not stained. This positive immunochemical reaction was not obtained in the same liver section when preimmune rabbit IgG was substituted for anti-HCV-NS5 IgG (Fig. 1b). Prior absorption of anti-HCV-NS5 IgG with its immunogen completely abolished specific immunostaining, as did omission of the primary antibody from the staining procedure (Fig. 1c). However, positive immunostaining of HCV-NS5 antigen was unchanged in sections pretreated with RNase (100 μg/ml) at 37° C for 20 min (Fig. 1d). HCV RNA was also clearly stained by *in situ* hybridization in the cytoplasm of hepatocytes but not in other organelles or in other types of liver cells (Fig. 2a). However, positive staining of HCV RNA was not found in sections pretreated with RNase (Fig. 2b).

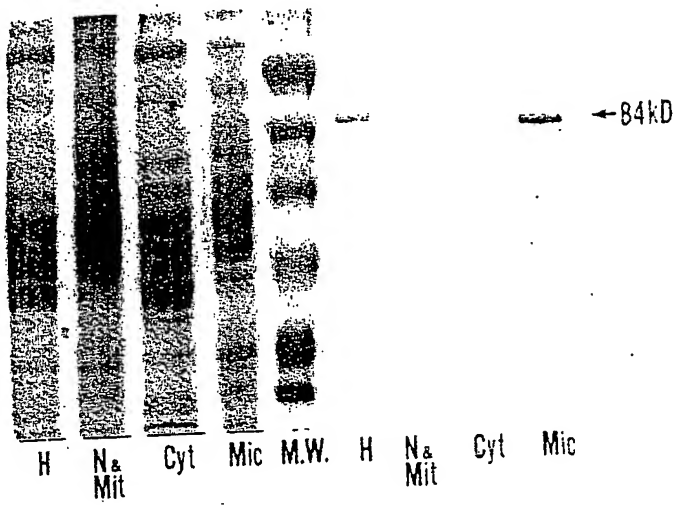



FIG. 3. SDS-PAGE and Western-blot analysis of human liver tissue with HCV-NS5 antigen. A single band at 84 kD was observed in both the total homogenates and the mitochondrial fractions. *H* = liver homogenate; *N & Mit*, nuclear and mitochondrial fraction; *Cyt*, cytosolic fraction; *Mic*, microsomal fraction; *M.W.*, standard proteins with molecular weights.

Western-blot analysis with HCV-NS5 IgG after SDS-PAGE of HCV marker–positive human liver samples is shown in Figure 3. A single band with a molecular weight of 84 kD was observed in both the total homogenates and the microsomal fractions but not in the nuclear and mitochondrial fractions or in the cytosol. No bands were detected by Western-blot analysis of peripheral blood mononuclear cells.

HCV-NS5 antigen was stained only in liver sections from HCV-positive patients and not in sections from patients with chronic type B hepatitis or alcoholic fibrosis. HCV-NS5 antigen was detected in 19 of 34 HCV marker–positive patients (56%). The detection rate was 33% in CPH, 52% in CAH and 86% in cirrhosis. HCV RNA was also stained by *in situ* hybridization in liver sections only from HCV-positive patients. HCV RNA was detected in 23 of 27 HCV marker–positive patients (85%), indicating a significantly higher detection rate than that of HCV-NS5 antigen by the $\chi^2$ test. The detection rate of HCV RNA was 67% in CPH, 86% in CAH and 100% in cirrhosis (Table 1).

HCV-NS5 antigen was stained in the hepatocytes of all acinus zones. Its location was not related to round cell infiltration. The staining pattern in hepatocytes was one of three types: diffuse, clustered or patchy. In the diffuse pattern, HCV-NS5 antigen was stained throughout the liver sections (Fig. 4a). In the cluster pattern, groups of hepatocytes in some parts of the section were stained strongly (Fig. 4b). In the patchy pattern, only a few isolated hepatocytes were stained (Fig. 4c). The staining patterns of HCV RNA by *in situ* hybridization in liver sections obtained from the same patients, as shown in Figure 4d, e and f, were quite similar to those of HCV-NS5 antigen in the corresponding liver sections (Fig. 4a, b and c). In two cases of CPH, which were HCV-NS5 antigen–positive, the staining pattern was patchy. In 11 cases of CAH, the patchy pattern was seen in two cases, the cluster pattern in seven and the diffuse pattern in one. In six cases of cirrhosis, the cluster pattern was found in two and the diffuse pattern in four. Distribution of the staining patterns of HCV RNA in each type of HCV-related disease was similar to that of HCV-NS5 antigen, although the diffuse pattern tended to be more frequent for HCV RNA (Table 2).

As determined on immunoelectron microscopy, HCV-NS5 antigen was stained in the cytoplasm of some but not all hepatocytes (Fig. 5a). At high magnification, HCV-NS5 antigen was stained in a fine granular fashion at the site of the endoplasmic reticulum (Fig. 5b). However, the other hepatocytic organelles, including the mitochondria and nucleus, were not stained. The glycogen areas of the hepatocytes were also negative (Fig. 5c). There was no staining in the Disse space and in nonhepatocytic cells. In HCV-negative cases, HCV-NS5 antigen was not stained in any part of the hepatocytes, except for an endogenous peroxidase reaction in the lysosomes (Fig. 5d).

In four of seven patients who were treated with interferon for 6 mo, HCV RNA in serum became negative at the end of the treatment. However, HCV-NS5 antigen and HCV RNA in liver sections were still positive in one patient but were negative in three patients. In all three patients who were positive for HCV RNA in serum at the end of interferon treatment, HCV-NS5 antigen and HCV RNA in liver sections were positive (Table 3).

## DISCUSSION

HCV is a positive single-strand RNA virus distantly related to flaviviruses (1, 13). Therefore RNA replicase, an RNA-dependent RNA polymerase, may be necessary for the replication of HCV, as well as other RNA viruses. RNA replicase has the motif characterized by a Tyr-Gly-Asp-(Thr)-Asp, which may be important in polymerase function by direct action in catalysis or by binding magnesium (5). In this study we synthesized the recombinant polypeptide with a 576-bp cDNA encoding a part of the NS5 region of the HCV genome, which has the Gly-Asp-Asp motif. Therefore an antibody against this recombinant polypeptide (anti-HCV-NS5 IgG) may recognize HCV polymerase–related antigen.

In Western-blot analysis with anti-HCV-NS5 IgG, an 84-kD protein was clearly detected as a single band in the microsomal fraction but not in the nuclear, mitochondrial or cytosol fractions. These results indicate that HCV-NS5 antigen, which may include HCV polymerase, is present in hepatocytes and may be produced in the rough endoplasmic reticulum. Immunohistochemically, HCV-NS5 antigen was revealed in the cytoplasm of hepatocytes but not in the nucleus or cell membrane. Moreover, as determined on immunoelectron microscopy, HCV-NS5 antigen was stained in fine granular fashion along the endoplasmic reticulum but not in other organelles. These results are compatible with the

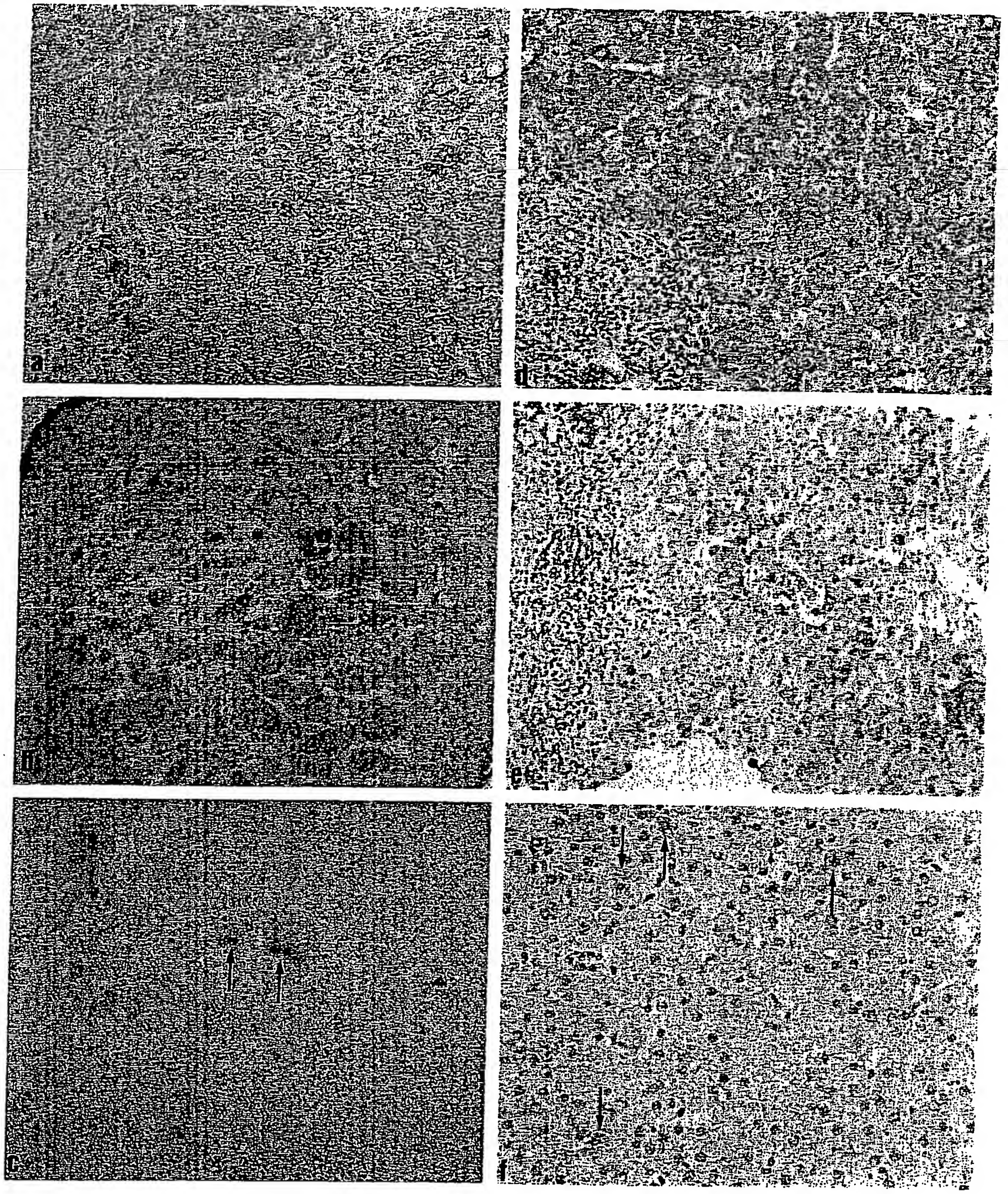


FIG. 4. The staining pattern of HCV-NS5 antigen and HCV RNA in liver sections of HCV-positive patients (×200). The immunostaining pattern of HCV-NS5 antigen was separated into three types: diffuse (a), cluster (b) and patchy (c). The staining pattern of HCV RNA in liver sections that were obtained from the same patients showing (a), (b) and (c), respectively, was similar to that of HCV-NS5 antigen in the corresponding liver section (d-f). Arrows denote positive cells.

hypothesis that RNA replicase is usually formed just after the viral RNA enters the cell and attaches to host ribosomes (2).

Grakoui et al. (15) reported that two proteins were derived from the HCV-NS5 region: NS5A (58 kD) and C-terminal NS5B (66 to 68 kD), when a cDNA encompassing the long open reading frame of HCV was used in vaccinia virus transient-expression assay. The NS5B protein was predicted to contain the RNA-dependent RNA polymerase activity on the basis of the presence of the characteristic Gly-Asp-Asp and surrounding conserved motifs. Although bacterially expressed HCV-NS5 peptide fragment was used for a part of NS5B protein in this study, the molecular size of HCV-NS5–related

**TABLE 2. Staining patterns of HCV-NS5 antigen and HCV RNA in liver sections**

| HCV markers in liver sections | Cases | Staining patterns (%) | | |
|---|---|---|---|---|
| | | Patchy | Cluster | Diffuse |
| HCV-NS5 antigen by immunostaining | | | | |
| CPH | 2 | 2 (100) | 0 (0) | 0 (0) |
| CAH | 11 | 3 (27) | 7 (64) | 1 (9) |
| Cirrhosis | 6 | 0 (0) | 2 (33) | 4 (67) |
| TOTAL | 19 | 5 (26) | 9 (47) | 5 (26) |
| HCV RNA by *in situ* hybridization | | | | |
| CPH | 2 | 1 (50) | 1 (50) | 0 (0) |
| CAH | 18 | 0 (0) | 10 (56) | 8 (44) |
| Cirrhosis | 3 | 0 (0) | 0 (0) | 3 (100) |
| TOTAL | 23 | 1 (4) | 11 (48) | 11 (48) |

**TABLE 3. Detection rate of HCV-NS5 and HCV RNA in liver sections from patients treated with interferon for 6 months**

| Serum HCV RNA | Cases | Liver sections (%) | |
|---|---|---|---|
| | | HCV-NS5 antigen (+) | HCV RNA (+) |
| Positive | 3 | 3/3 (100) | 3/3 (100) |
| Negative | 4 | 1/4 (25) | 1/4 (25) |

antigen detected in human liver was 86 kD and thus slightly larger than that of NS5B. This discrepancy may have resulted in different host cells, which were cultured mammalian cells in Grakoui's study (15) and were human liver cells in this study. Furthermore, we observed the products derived from native HCV, whereas Grakoui et al. (15) observed the polypeptide expressed from HCV cDNA in vaccinia virus.

Recently, Takehara et al. (16) reported that when an RT-PCR technique was used, the minus strand of HCV RNA was only detected in the liver and not in the plasma or in peripheral-blood mononuclear cells, which suggested that HCV replicates in the liver but not in the peripheral-blood mononuclear cells. In this study, HCV-NS5 antigen and HCV RNA were stained only in hepatocytes and not in other types of liver cells, including infiltrating lymphocytes. It was absent from peripheral blood mononuclear cells (data not shown). These results suggest that HCV may be replicated only in hepatocytes and not in hepatic or peripheral-blood mononuclear cells. On the other hand, Lamas et al. (17) reported that the minus strand of HCV RNA was detected in both hepatocytes and infiltrating mononuclear cells by *in situ* hybridization and suggested that HCV replication may occur in both cells. However, they studied only four needle biopsy specimens from patients who were coinfected with human immunodeficiency virus and HCV; the signal for sense HCV RNA probes was detected in only very few cells, and its intensity was very low even in hepatocytes. It is therefore very difficult to determine whether infiltrating mononuclear cells contain the minus strands of HCV RNA.

The staining reaction for HCV-NS5 antigen was not abolished by RNase pretreatment, indicating that stained HCV-NS5 antigen is not a virus itself. However, the staining pattern for HCV-NS5 antigen was quite similar to that for HCV itself detected by *in situ* hybridization with the use of a cDNA probe for the core region, indicating that HCV-NS5 antigen is stained in hepatocytes infected by HCV.

HCV-NS5 antigen was stained only in liver sections from HCV-positive patients but not in any sections from patients with chronic type B hepatitis or with alcoholic fibrosis. In chronic type C liver diseases, the detection rate of HCV-NS5 antigen in liver sections was higher in cirrhosis than in CPH or CAH. This supports the hypothesis that replication may occur more frequently in the advanced stages of HCV-related liver diseases.

Hiramatsu et al. (18) reported that in anti-C100-3–positive patients, positive immunostaining with antibodies for the core, envelope and NS3 regions of the HCV genome in liver biopsy specimens was found in 23%, 24% and 24%, respectively. Compared with their detection rates, the present anti-HCV-NS5 IgG detection system was more sensitive. Hosoda et al. (19) reported that the detection rate of HCV RNA in liver tissues with the RT-PCR technique was 63% in patients with chronic non-A, non-B hepatitis and 71% in anti-C100-3–positive patients, rates that are roughly similar to the overall detection rate of HCV-NS5 antigen in this study. Recently, interferon therapy has been used for chronic type C hepatitis; however, it is very difficult to determine when HCV is eliminated from the liver, even if HCV RNA is not detected in serum. In this study HCV-NS5 antigen in a liver section was still positive in a patient whose serum HCV RNA became negative with interferon treatment. Although the overall detection rate of HCV-NS5 antigen is not high, immunostaining of HCV-NS5 antigen in liver biopsy sections is useful in

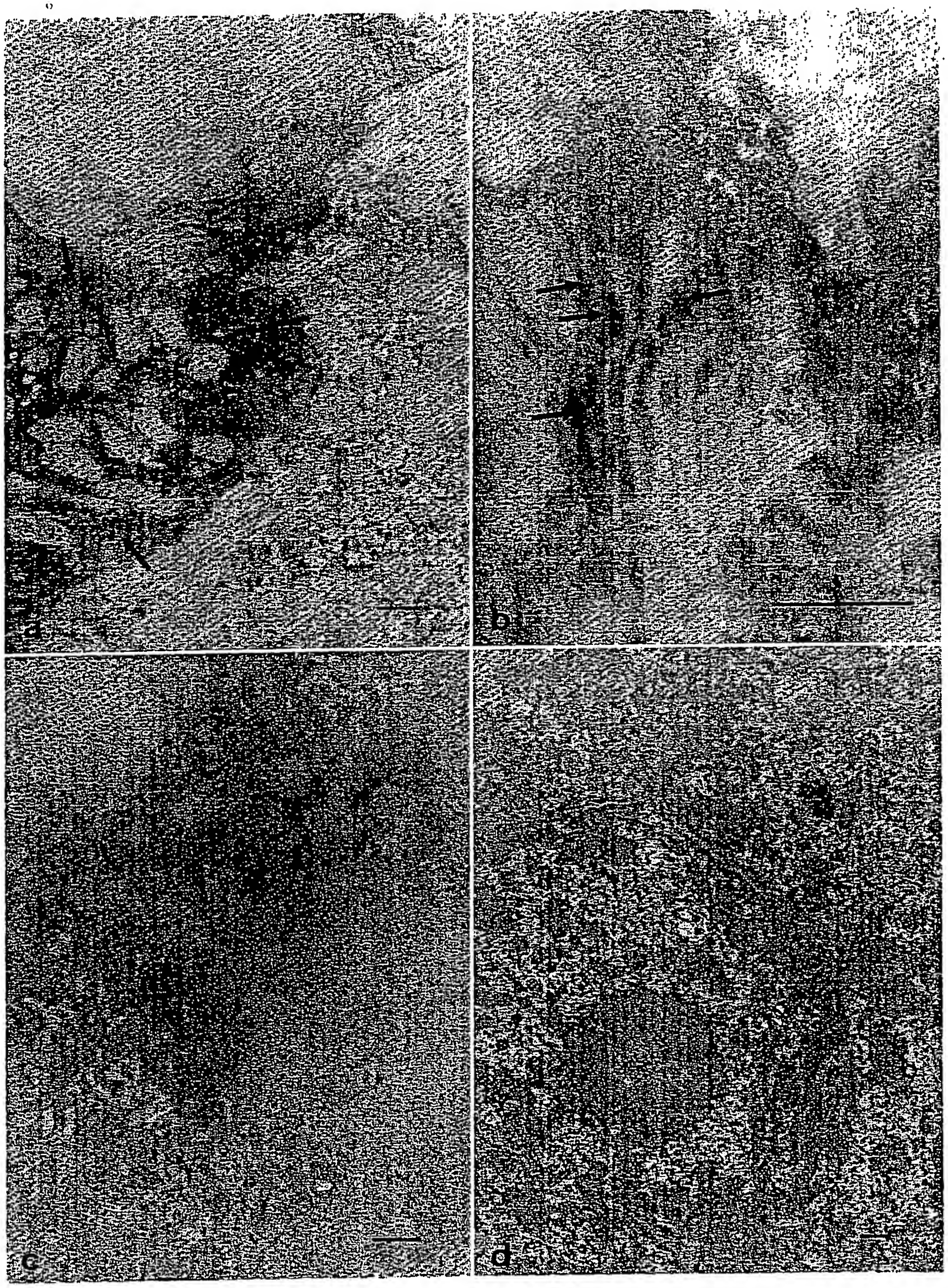


FIG. 5. Immunoelectron microscopic findings of HCV-NS5 antigen in liver section. (a) HCV-NS5 antigen was stained in cytoplasm, as indicated by arrows. (b) At high magnification, HCV-NS5 antigen was stained in fine granular fashion together with endoplasmic reticulum as indicated by arrows. (c) The staining reactions were not observed in the glycogen areas of hepatocytes. (d) In HCV-negative patients, HCV-NS5 antigen was not detected in hepatocytes. The positive reaction in the lysosomes is due to endogenous peroxidase.

determining the effect of interferon therapy and evaluating the cessation of HCV replication when paired liver sections before and after treatment were available. The results of this study indicate that the detection of HCV RNA by *in situ* hybridization is more sensitive for this purpose. Detection of HCV RNA itself in liver tissues with the RT-PCR technique may also be more sensitive (16, 19). However, in comparison with immunostaining of HCV-NS5 antigen, these two sophisticated methods are extremely complicated.

*Acknowledgment:* We are grateful to Professor Peter J. Scheuer (University of London) for help in editing this manuscript.

## REFERENCES

1. Choo QL, Kuo G, Weiner AJ, Overby LR, Bradley DW, Houghton M. Isolation of a cDNA clone derived from a blood-borne non-A, non-B viral hepatitis genome. Science 1989;244:359-362.
2. Watson JD, Hopkins NH, Roberts JW, Steutz JA, Weiner AM. The replication of bacterial viruses. In: Watson JD, Hopkins NH, Roberts JW, Steutz JA, Weiner AM, eds. Molecular biology of the gene. Menlo Park, CA: The Benjamin/Cummings Publishing Company, Inc., 1987;1:503-543.
3. Enomoto N, Takase S, Takada A, Date T. Detection of hepatitis C virus genomes from patient's plasma using PCR method. Gastroenterol Jpn 1990;25:404.
4. Enomoto N, Takada A, Nakao T, Date T. There are two major types of hepatitis C virus in Japan. Biochem Biophys Res Commun 1990;170:1021-1025.
5. Argos P. A sequence motif in many polymerases. Nucleic Acids Res 1988;16:9909-9916.
6. Mimms L, Vallari D, Ducharme L, Holland P, Kuramoto IK, Zeldis J. Specificity of anti-HCV ELISA assessed by reactivity to three immunodominant HCV regions. Lancet 1990;336:1590-1591.
7. Okamoto H, Okada S, Sugiyama Y, Tanaka T, Sugai Y, Akahane Y, Machida A, et al. Detection of hepatitis C virus RNA by a two-stage polymerase chain reaction with two pairs of primers deduced from the 5'-noncoding region. Jpn J Exp Med 1990;60: 215-222.
8. Enomoto N, Nakao T, Takada A, Date T. Genotype of hepatitis C virus in Japan. Jpn J Clin Med 1991;49:314-318 (in Japanese).
9. McKinney MM, Parkinson A. A simple, non-chromatographic procedure to purify immunoglobulin from serum and ascites fluid. J Immunol Methods 1987;96:271-278.
10. Schenkman JB, Cinti DL. Preparation of microsomes with calcium. Methods Enzymol 1978;53:83-89.
11. Laemmli UK. Cleavage of structural proteins during the assembly of the head of bacteriophage T4. Nature 1970;227:680-685.
12. Nielsen PJ, Manchester KL, Towbin H, Gordon J, Thomas G. The phosphorylation of ribosomal protein S6 in rat tissues following cyclohexisamide injection, in diabetes, and after denervation of diaphragm., J Biol Chem 1982;257:12316-12321.
13. Miller RH, Purcell RH. Hepatitis C virus shares amino acid sequence similarity with pestiviruses and flaviviruses as well as members of two plant virus supergroups. Proc Natl Acad Sci U S A 1990;87:2057-2061.
14. Tanaka Y, Enomoto N, Kojima S, Tang L, Goto M, Marumo F, Sato C. Detection of hepatitis C virus RNA in the liver by in situ hybridization. Liver 1993;13:203-208.
15. Grakoui A, Wychowsky C, Lin C, Feinstone SM, Rice CM. Expression and identification of hepatitis C virus polyprotein cleavage products. J Virol 1993;67:1385-1395.
16. Takehara T, Hayashi N, Mita E, Hagiwara H, Ueda K, Katayama K, Kasahara A, et al. Detection of the minus strand of hepatitis C virus RNA by reverse transcription and polymerase chain reaction: implications for hepatitis C virus replication in infected tissue. HEPATOLOGY 1992;15:387-390.
17. Lamas E, Baccarini P, Housset C, Kremsdorf D, Brechot C. Detection of hepatitis C virus (HCV) RNA sequences in liver tissue by in situ hybridization. J Hepatol 1992;16:219-223.
18. Hiramatsu N, Hayashi N, Haruna Y, Kasahara A, Fusamoto H, Mori C, Fuke I, et al. Immunohistochemical detection of hepatitis C virus-infected hepatocytes in chronic liver disease with monoclonal antibodies to core, envelope and NS3 regions of the hepatitis C virus genome. HEPATOLOGY 1992;16:306-311.
19. Hosoda K, Yokosuka O, Omata M, Kato N, Ohto M. Detection and partial sequencing of hepatitis C virus RNA in the liver. Gastroenterology 1991;101:766-771.